DOCUMENT RESUME

ED 127 368 95 TM 005 528

AUTHOR Sandoval, Jonathan
TITLE Beginning Teacher Evaluation Study: Phase II, 1973-74, Final Report: Volume III.3. The Evaluation of Teacher Behavior Through Observation of Videotape Recordings.
INSTITUTION California State Commission for Teacher Preparation and Licensing, Sacramento.; Educational Testing Service, Princeton, N.J.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO ETS-PR-76-10
PUB DATE Jul 76
NOTE 127p.; For related documents, see TM 005 524-535

EDRS PRICE MF-$0.83 HC-$7.35 Plus Postage.
DESCRIPTORS Academic Achievement; *Classroom Observation Techniques; *Effective Teaching; Elementary Education; Grade 2; Grade 5; Reliability; *Student Behavior; *Teacher Behavior; Validity; *Video Tape Recordings

ABSTRACT

The Beginning Teacher Evaluation Study (BTES), Phase II, was a research project on effective teaching behavior--what teachers do that significantly affects what and how pupils learn. The purposes of Phase II were to (1) develop an assessment system for measuring teacher and pupil behaviors and other factors which could influence each of them and their interrelationships and (2) generate hypotheses about the interrelationships among teacher and pupil behaviors and related factors. Forty-one second grade and 54 fifth grade experienced teachers participated in the study. This volume of the final report is based on the videotaped recordings of 94 teachers taking part in the BTES study. The investigation was directed at providing tentative answers to the following questions about the utility of videorecording: Do the advantages outweigh the disadvantages in yielding reliable and valid information about teacher and student behavior in the classroom? What kind of information can be obtained by observers viewing the videotape? Can different kinds of teaching behavior and student behavior be observed? Can two or more viewers obtain the same kind of information from separate viewings of the videotape? How consistent are the observed teacher behaviors? And, most important, once a set of reliable and stable teacher behaviors have been isolated, do these teacher behaviors have anything to do with pupil achievement in the classroom? (RC)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED127368

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
DOROTHY URBAN
ETS
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

TM005 528

PR-76-10

BEGINNING TEACHER EVALUATION STUDY
PHASE II 1973-74

FINAL REPORT

VOLUME III.3.

THE EVALUATION OF TEACHER BEHAVIOR THROUGH OBSERVATION OF VIDEOTAPE RECORDINGS

BY

JONATHAN SANDOVAL
UNIVERSITY OF CALIFORNIA
DAVIS

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

BEGINNING TEACHER EVALUATION STUDY: PHASE II

FINAL REPORT: VOLUME III.3.

THE EVALUATION OF TEACHER BEHAVIOR THROUGH OBSERVATION OF VIDEOTAPE RECORDINGS

by

Jonathan Sandoval

University of California
Davis, California

A project conducted by Educational Testing Service for the California Commission for Teacher Preparation and Licensing and funded by the National Institute of Education.

Copyright © 1975, 1976 by Educational Testing Service. All rights reserved.

PREFACE

The Beginning Teacher Evaluation Study (BTES) is a long-term project of the California Commission for Teacher Preparation and Licensing. The Commission is responsible for licensing teachers in California and is trying to determine what factors should be considered in this process.

The second phase of the study was conducted by Educational Testing Service for the Commission. Phase II was the hypotheses-generating and instrument-development phase of BTES. ETS had two tasks: (1) to develop an assessment system to measure both teacher and pupil behaviors as well as other factors which might be related to these behaviors; and (2) to generate hypotheses about the interrelationships between teacher and pupil behaviors and related factors.

The study was conducted in 43 schools in eight districts throughout the state of California. A total of 41 second grade teachers and 54 fifth grade teachers participated in the project during Phase II.

The final report for Phase II consists of several volumes. Volume I describes the design and rationale for the experimental design and data analysis procedures and includes the major findings of Phase II. Volume II describes the conduct of the field study and the sample of participants.

Because of the complex nature of Phase II, a variety of techniques was used to measure teacher and pupil behaviors. They are described in Volumes III, IV, and V. Results are also included in these volumes.

Volume III describes the observation systems in detail and is available in three separately bound sections. The first section, Volume III.1., describes the behavior recording observation system used in the project--APPLE (Anectdotal Process for Promoting the Learning Experience). Volume III.2. describes the category system used to observe classroom activities--

RAMOS (Reading and Mathematics Observation System). The third section of this volume, III.3., covers the videotaping of instructional activities during reading and mathematics.

Volume IV concerns other aspects of the measurement system and covers both the pupil and teacher test batteries.

The fifth volume covers a series of small studies done as part of Phase II. Volume V.1. looks at teacher aptitudes as related to teacher behaviors. Volume V.2. is concerned with the relationship between teacher expectations and pupil performance. Volume V.3. reviews performance of pupils in the BTES teachers' classrooms for two years prior to Phase II, the historical test data. Volume V.4. discusses the Diagnostic Film Test, a device designed to assess teachers' skills in diagnosing reading problems and prescribing corrective action. Volume V.5. summarizes the results of work diaries completed by the teachers on their reading and mathematics instructional program.

Information on the availability of these volumes can be obtained from:

Dr. Frederick J. McDonald
Educational Studies
Educational Testing Service
Princeton, NJ 08540

Information on other phases of BTES can be obtained from:

California Commission for Teacher
Preparation and Licensing
1020 O Street
Sacramento, CA 95814

ACKNOWLEDGEMENTS

The author acknowledges the work of the Davis APPLE recorders, Carol Branan, Judy Dudek, Jim Pagliuso, Dwayne Ramsey and Lee Wilce. They worked long and diligently. This project would not have been successful without their many contributions. They warrant special thanks and recognition.

Jonathan Sandoval

TABLE OF CONTENTS

	Page
Preface	i
Acknowledgements	iii
The Evaluation of Teacher Behavior Through Observation of Videotape Recordings	1
Sample	2
Procedure	3
Instruments	6
Observation systems	6
Rating Schedules	8
Application of Instruments	10
Observer Training	10
Coding Procedures	12
Coding with APPLE	12
Coding with Interaction Analysis	12
Subsample Inter-Rater Reliability Procedures	13
APPLE	13
Interaction Analysis	15
Rating scales	15
Results of Subsample Inter-Rater Reliability	17
APPLE	20
Interaction Analysis	29
Rating scales	30
The Consistency of Subsample Teacher Behavior Over Time	37
Procedures	37

TABLE OF CONTENTS (continued)

Page

Results of Consistency of Subsample Teacher Behavior Over Time 38

APPLE 38

Interaction Analysis 39

Rating scales 39

Frequency of Occurrence of Events in BTES Classrooms 47

Teacher Instruction Lexicon 47

Teacher Feedback Lexicon 51

Data Reduction 61

Teacher Instruction events 61

Teacher Feedback events 61

Rating scales 63

Relationship Between Information Collected from Videotape and Pupil Achievement 68

Teacher Feedback Lexicon 71

Recommendations 88

Videotape recording 88

Videotape coding 89

Hypotheses for further research 90

References 93

Appendix A 95

Appendix B 101

Appendix C 111

LIST OF TABLES

Table		Page
A	Descriptive Information on 12 Subsample Teachers	4
1	Inter-rater Reliability of Two Recorders Coding Five Teachers During Reading and Mathematics Lessons on Three Occasions Using Teacher Instruction Lexicon . .	22
2	Inter-rater Reliability of Three Recorders Coding Five Teachers During Reading and Mathematics Lessons on Three Occasions Using Teacher Feedback Lexicon . . .	25
3	Inter-rater Reliability of Four Recorders Coding Six Teachers on Time Three Tapes with Interaction Analysis	32
4	Average Inter-Rater Reliabilities Across Six Teachers with Interaction Analysis Calculated by Scott's Pi . . .	33
5	Inter-Rater Reliability of Four Recorders Rating Six Teachers on Time Three Tapes with Videotape Rating Schedule #1	34
6	Inter-rater Reliability of Three Recorders Rating Five Teachers on Three Tapings of Reading and Mathematics with Videotape Rating Schedule #2	35
7	Inter-rater Reliability of Two Recorders Rating Five Teachers on Three Tapings of Mathematics and Reading Lessons Using Observation Classroom Summary Record: APPLE System	36
8	Consistency Over Three Tapings of Ten Teachers on Eleven Events From APPLE Teacher Instruction Lexicon and Teacher Feedback Lexicon	42
9	Consistency Over Two Tapings of 12 Teachers on 10 Categories of Interaction Analysis	43
10	Consistency Over Three Tapings of 12 Teachers on Six Items from Videotape Rating Schedule #1 with High Inter-Rater Reliability	44
11	Consistency Over Three Tapings of 10 Teachers on Videotape Rating Schedule #2	45
12	Consistency Over Three Tapings of 10 Teachers on Selected Items of Observation Summary Record: APPLE System	46

LIST OF TABLES (continued)

Table		Page
13	Mean Observed Frequencies of Teacher Instruction Categories: Total Group, Reading/Math Subgroups, Grade Level Subgroups	49
14	Mean Observed Frequencies of Teacher Instruction Categories: Socioeconomic Subgroups	50
15	Mean Observed Frequencies of Teacher Feedback Categories for Achievement: Total Sample and Grade Level Subgroups	53
16	Mean Observed Frequencies of Teacher Feedback Categories for Achievement: Reading and Math Subgroups	54
17	Mean Observed Frequencies of Teacher Feedback Categories for Achievement: Socioeconomic Subgroups	55
18	Mean Observed Frequencies of Teacher Feedback Categories for Behavior: Total Sample and Grade Level Subgroups	58
19	Mean Observed Frequencies of Teacher Feedback Categories for Behavior: Reading and Math Subgroups	59
20	Mean Observed Frequencies of Teacher Feedback Categories for Behavior: Socioeconomic Subgroups	60
21	Combined Categories for the Teacher Instruction Lexicon	64
22	Combined Categories for the Teacher Feedback Lexicon	65
23	Factor Structure Matrix for Reliable Observation Summary Record Items	66
24	Correlation Coefficients Between Teacher Instruction Behavior Events and Spring Test Scores Partialling Out Fall Test Scores	67
25	Correlation Coefficients Between Teacher Feedback Behavior Event Categories and Spring Test Scores Partialling Out Fall Test Scores	75

LIST OF TABLES (continued)

Table		Page
26	Correlation Coefficients Between Rating Scale Items and Factor Scores, and Spring Test Scores Partialling Out Fall Test Scores	80
27	Mean Ranks in Comparisons Between Top and Bottom Reading and Mathematics Teachers on Teacher Instruction Lexicon Events	85
28	Mean Ranks in Comparisons Between Top and Bottom Reading and Mathematics Teachers on Teacher Feedback Lexicon Events	86
29	Mean Ranks in Comparisons Between Top and Bottom Reading and Mathematics Teachers on Rating Scales . . .	87

LIST OF FIGURES

Figure		Page
1	General reliability design for APPLE	14
2	Reliability design for Interaction Analysis	16
3	General design for studying consistency	41

The Evaluation of Teacher Behavior Through Observation of Videotape Recordings

For many years, the videotape recording of teachers teaching has proven valuable as a means of giving teachers feedback to improve their classroom effectiveness. Fuller and Manning (1973) have prepared a comprehensive review of this use of videorecordings. Techniques ranging from structured microteaching to the informal self recording by teachers have been encouraged by teacher education institutions and school district staffs. One of the oldest educational tools for helping people improve performance is, after all, the mirror. Modern technology has provided us with a mirror-with-a-memory. It is little wonder that videotaping has been enthusiastically received. The equipment necessary for videotaping is relatively inexpensive and can be operated effectively after a small amount of training. During this hypothesis generating year, one of the aims of the BTES project has been to evaluate videotaping as a means of collecting information on teaching behavior, information that can be related to pupil outcome data.

As a means of collecting research data, videotaping has some advantages over live observation in classrooms. Because the teacher's behaviors have been preserved on tape, they may be examined over and over by one or more observers. The tape may be stopped and started any number of times permitting the observers to make more considered and deliberate decisions about what is occurring and to examine behavior in close detail. Videotaping has a number of disadvantages, however. Having a camera and operator in the classroom can be very artificial and distracting.

In addition, the behaviors recorded represent only a fraction of those which may be taking place in the classroom because a camera has tunnel vision, is not mobile, and the medium does not allow for a very clear view of its subject.

The major question that must be asked about videotaping is whether the advantages outweigh the disadvantages in yielding reliable and valid information about the behaviors of teachers in classrooms. The first problem that must be addressed is what kind of information can be obtained by observers viewing the videotapes. Can different kinds of teaching behavior be observed and can different kinds of pupil behavior be observed? The second problem is whether or not two or more observers collect the same kinds of information from separate viewings of the videotapes. The third problem is how consistent are the observed teacher behaviors. Do they represent stable and lasting aspects of the teacher's style in the classroom? The final and most important problem to be examined is, once a set of reliable and stable behaviors has been isolated, do these teacher behaviors have anything to do with pupil achievement in the classroom.

This investigation was directed at providing tentative answers to these questions about the utility of videorecording. The report is organized in a serial manner dealing with each problem in turn. First, however, the procedures followed in accomplishing the task are detailed.

Sample

Project videotapers recorded the teaching of ninety-four teachers taking part in the BTES study. (See Volume II.)

In general, the teachers were volunteers but came from a wide variety of schools. They were all experienced with an average 13.4 years teaching. One participating teacher, the 95th, was not videotaped because of illness. Most were taped three times on one visit, but a subsample of 12 teachers was taped on three or four separate occasions. The purpose of the subsample was to provide information about the consistency of observed behavior over time. Table A contains the characteristics of this special subsample.

Procedure

Details on the videotape equipment, the training of operators, and the scheduling are to be found in Volume II. For each visit to a classroom, where possible, the videotapers followed the same general procedure. One tape was made of the opening exercises at the start of the day. The intent of this first taping was to acclimate the pupils to the novelty of the videotaping and to desensitize the teacher. After this warm-up tape, which was not used in the data analysis, the videotapers recorded both the reading and mathematics lessons scheduled for the day. In general, the camera operator focused on the teacher if the teacher presented the lesson to the entire class. As much as possible, the closeup shots of the teacher were recorded, but from time to time the range was widened to include the entire class. When a teacher worked with individuals or small groups, the camera was focused on as many of the significant interactors as possible. The camera was mounted in one location, and the operator did not move during a taping.

Table A

Descriptive Information on 12 Subsample Teachers

Teacher Number	Sex	Years Teaching	Grade Taught	School Socioeconomic Class	District Size	District Urbanization	Average Student Achievement in Reading (Intercepts Compared to Other Classrooms in Project)	Average Student Achievement in Mathematics (Intercepts Compared to Other Classrooms in Project)	Instrument with Which Observed to Test Reliability
1	M	22	5	Middle	Medium	Rural	2nd 25%	Bottom 25%	TF, VRS, TI, OSR, I.A.
2	F	8	5	Lower Middle	Medium	Rural	Bottom 25%	2nd 25%	TF, VQS, I.A.
3	M	21	5	Higher Middle	Small	Suburban	2nd 25%	Top 25%	TF, VRS, I.A.
4	F	5	5	Lower	Large	Urban	Top 25%	2nd 25%	TF, VRS
5	F	14	5	Lower	Large	Urban	--	3rd 25%	TF, VRS
6	F	20	2	Middle	Medium	Rural	3rd 25%	3rd 25%	TI, OSR
7	F	24	5	Middle	Medium	Rural	2nd 25%	3rd 25%	TI, OSR, I.A.
8	M	23	5	Higher Middle	Small	Suburban	Top 25%	Top 25%	TI, OSR
9	M	5	5	Middle	Small	Suburban	Bottom 25%	2nd 25%	TI, OSR
10	F	26	5	Higher Middle	Large	Urban	Top 25%	Top 25%	I.A., VRS #1
11	F	15	5	Higher Middle	Small	Suburban	Top 25%	--	I.A., VRS #1
12	F	28	2	Higher Middle	Small	Suburban	Top 25%	Second 25%	--

The first teachers recorded were members of the subsample. The first taping became, in essence, a trial run. One problem emerged and was solved at this time. On the first tapes from the 12 subsample teachers, viewers of the finished tapes reported that the teacher's speech could not be clearly understood when there was a considerable amount of background noise. During the taping different kinds of microphones were tried out to determine the most effective way of picking up what both the pupil and the teacher said during the lesson. It became clear that a portable cordless lavaliere microphone and transmitter were best; this equipment was employed during the balance of the videotapings.

The procedure for the taping of the subsample was different from that of the total sample in one respect. For the second and third tapings two cameras and tape recorders were used. The operator followed the strategy described above for one camera while using the other camera to sweep over the class, not focusing on the teacher. Those who viewed the completed tapes decided that this second camera was of limited usefulness because it often pointed the wrong direction to record a pupil response, and because individual pupils and materials could not be distinguished on a TV monitor. The procedure of the two cameras was not used with the total sample.

Because the project staff wished to give the cooperating teachers a chance to view themselves on the tape, provisions were established to lend the tapes to the teachers. If a teacher wished to see the tape of the opening exercise, it was turned over on the spot on the day of the taping. If a teacher desired to view the

reading or mathematics tape, the tape was returned by mail after it had been coded by the project observers.

Generally, the videorecording went well and the tapes were of acceptable quality. Some of the tapes have sections where there were video breakdowns, and some of the tapes have sections missing sound; but almost all could be coded. One mathematics tape for one of the 94 teachers and one reading tape for a different teacher were too poor to be used. The project teachers were most cooperative with the taping. All but one were videotaped and only two of the 12 subsample teachers declined to be videotaped on a fourth visit to their classrooms.

All of the teachers were recorded during March and April of the year. The subsample was taped additionally in early December, in late January-early February, and in the second half of May.

Instruments

In addition to hypotheses generation, the major thrust of this BTES year has been instrument development. Several instruments were created or adapted for use in analyzing the videotape recordings. They fall into the general categories of observation systems and rating schedules.

Observation systems. Both APPLE and RAMOS personnel viewed videotapes recorded in the subsample classrooms. They were asked to make a subjective evaluation of whether or not they could use each of their systems in coding these videotapes. The RAMOS staff felt that it might be possible to code the tapes with no modifications in their system. Accordingly, each of the subsample videotapes was

coded by a RAMOS observer.*

The staff felt that although, as presently constituted, the APPLE system could be used to process the information on the tapes, the richness of the data would be sharply reduced compared with those collected in "live" classrooms. The fidelity of the videotape medium did not permit an observer to recognize individual children, or to see many of the details of the interactions between the teacher and a pupil. The observation rating form might be used, but not the standard event lexicon. The staff decided to develop two new APPLE lexicons for the BTES project, the Teacher Instruction Lexicon and the Teacher Feedback Lexicon, which are devoted primarily to the observation of teaching behaviors. These lexicons seemed to be more adaptable to the videotapes because of their focus on the teacher. In the interest of time, they also decided to omit the collection of detailed anecdotes and to allow the observers simply to record the event names. Appendices A and B contain a list of the events in the two lexicons along with the event definitions.

The Teacher Instruction Lexicon is a listing of the various categories of teacher behaviors used in conducting a lesson to one or more children in a classroom. The numbering in the lexicon is more or less chronological for a lesson. The first three categories are all different kinds of introductory remarks. The next eight categories are different kinds of explanations or expository remarks a teacher might make in the course of presenting information to be

---

*The data were collected for RAMOS, but it was determined that there was too much missing information for it to be analyzed properly. There were not enough instances of RAMOS categories to code. It was decided to drop RAMOS as a method for analyzing behavior on videotapes of the larger sample.

learned. The last nine categories are a set of questioning or testing behaviors a teacher might use to stimulate mastery learning in pupils, or to evaluate what learning has taken place.

The Teacher Feedback Lexicon is an exhaustive listing of the ways in which teachers report information back to pupils about the pupils' performance in learning or in following classroom rules. The numbering system in this lexicon has some logic to it. Across the three general categories of (1) Positive Responses, (2) Non-Committal Responses and (3) Negative Responses, the letters in the code are parallel. "A" responses are simple informational ones, "B" responses are elaborations, "C" responses are "socratic" reactions, "D" responses are praise or punishment acts, "E" responses are attempts to create peer involvement, and "F" responses involve ignoring. In recording the behaviors listed as teacher feedback, the observers were to note whether or not the feedback was directed at arranging pupil's achievement or behavior in the classroom.

Because the APPLE lexicons were new and needed to be validated, we decided to employ a standard observation system which could, presumably, collect reliable and useful information from the tapes. We selected Flander's Interaction Analysis (I.A.) because it is a teacher-focused observation tool, it has been used with videotapes in the past, and it is thoroughly documented. The plan was for Interaction Analysis to serve as a benchmark against which to evaluate the APPLE observation system's utility.

Rating schedules. Perhaps the easiest method of collecting data on teachers is to have observers rate certain aspects of the

teacher's performance in the classroom. Observers require relatively little training to prepare them for their task and, in rating, contribute global clinical insights that might not otherwise be made explicit through quantitative procedures. Since the APPLE observers had completed a rating schedule, "Observation Classroom Summary Record: APPLE System," in many of the classrooms following a live observation, we decided to ask the videotape observers to complete this form. Reproduced in Appendix C, this schedule calls for general judgements about the nature and amount of feedback teachers give pupils, about the "directiveness" of teachers, about the quality of pupil responses, about the sensitivity of the pupils to the videotaping, about the classroom organization, and about the general emotional climate in the classroom. A four-point scale, ranging from "not at all descriptive" to "very descriptive," accompanied each of the 24 items. Video observers filled out this questionnaire after they had coded both the reading and mathematics lessons using one of the APPLE lexicons.

Although the existing rating schedule did tap some important dimensions of what could be observed in the classroom, after some thought we generated new items which became Videotape Rating Schedule #1 and Videotape Rating Schedule #2 (See Appendix C). The items on these schedules are concerned with the pace, tempo and variety in the lesson, the teacher's movements and enthusiasm, the organization and emphasis of the lesson, the teacher's empathy for the students, the pupils' reaction to the lesson, the teacher's questioning strategy, and the Rogerian concepts of warmth and honesty. In this schedule, however, there are four possible ratings, each anchored to a specific behavioral description. The video observers judged the

19 items on these two schedules after coding each reading and mathematics tape with the APPLE Teacher Feedback Lexicon.

Application of Instruments

All of the instruments, both rating schedules and observation systems, were used in coding the videotapes recorded in the subsample classrooms. The first taping sound quality was such that the observers were comfortable only in completing the rating schedule. With the change in microphone for the second, third and fourth tapings, it was possible to code these tapes with each of the observation systems. As the I.A. system was not of primary interest, the observers did not apply it to the fourth taping. They did code the late May tapes with APPLE.

After collecting the data from the coding of the subsample, we found it possible to evaluate the rating items and the observation systems. We needed to decide how to most efficiently code the tapes from the remaining 82 classrooms. First we simplified the rating scales by throwing out those items which had questionable inter-rater reliability. We decided to use the two APPLE lexicons to code the major sample tapes.

Observer Training

The Davis observers, although they had not done APPLE classroom observation, came very well prepared for their task with the BTES tapes. The four principal observers were credentialed, experienced classroom teachers, had completed a graduate course in observing and improving college level teaching, and had been employed on the Davis campus in an inservice education project aimed at improving the

effectiveness of teaching assistants and other instructors. Their job for that project involved videotaping teachers in college classrooms, analyzing the tapes with Interaction Analysis and other systems, then sitting down with the instructor to review the taped lesson. After almost six months or more of this experience, they were quite familiar with videotapes of teachers teaching and I.A. After a short orientation to elementary grade classrooms and a few practice sessions, they started in on the BTES tapes. Throughout the time they coded the tapes, they were able to consult with an acknowledged expert on I.A., who teaches on the Davis campus. In the beginning they spent several hours with him going over tapes which were particularly difficult to code.

After they completed the I.A. portion of the coding, the same four observers became APPLE videotape observers. One other graduate student had been assigned to the APPLE lexicons and had engaged in working out training procedures while the others did the I.A. coding. She spent a month and a half working out training materials, transcripts and, in general, testing the APPLE videotape lexicons. She conducted most of the training sessions and coded tapes for the reliability-consistency study.

Before coding on their own, the observers first had to memorize the APPLE Lexicon they would use. Two observers learned the Teacher Feedback Lexicon and two observers mastered the Teacher Instruction Lexicon. After achieving a thorough grounding in the lexicons, each observer coded tapes in tandum with the experienced observer-trainer so that they could discuss divergent scoring. Next, each observer coded a tape that had been carefully coded previously, and compared

their results with the standard. When their coding matched well with the standard, the observers were allowed to start coding new tapes independently. Once the coders started on the tapes, they met as a group once or twice a week to clarify problems and inconsistencies. In addition, they met informally with each other and often consulted with the author as the need arose. They were free to discuss the coding as long as they did not compare notes on specific tapes. They were warned against any collusion which might artificially increase inter-rater reliability.

Coding Procedures

Coding with APPLE. The APPLE coder's task was to classify one of two kinds of teacher actions present on the tape, giving didactic instruction and giving feedback. Depending on the lexicon assigned, the observer watched the videotape, waiting until an instructional or a feedback behavior occurred. When a behavior occurred which fit a lexicon category, the observer wrote down the number of the category. There was no limit on the number of events to be noted, and there was no limit to the number of times the tape could be halted or replayed in making judgements. The observers' instructions were to record 40 minutes of each tape, or to the end of the lesson, whichever occurred first. After the completion of the coding of a particular tape, each observer completed one or more of the rating scales.

Coding with Interaction Analysis. The Interaction Analysis coding was quite straightforward. Following a very brief preview of the tape, each observer began the process of summing up each

three-second time period in the lesson with the I.A. category which best fit. Since a running record of the time in the lesson was not important, the I.A. observers were able to stop and start the recorder, and to replay sections of the lesson in order to make more accurate judgements. The I.A. coders were also instructed to record 40 minutes of each tape or the entire lesson if the lesson was less than 40 minutes long.

Subsample Inter-Rater Reliability Procedures

APPLE. Two observers coded the tapes with the APPLE Teacher Instruction Lexicon (TI) and three with the Teacher Feedback Lexicon (TF). The procedure outlined by Medley and Mitzel in the first Handbook of Research on Teaching (1963) was followed to generate an estimate of inter-rater reliability. Each observer coded the reading and mathematics tapes of five teachers on three of the tapings (second, third, and fourth tapings). See the last column of Table A for details on which teachers were coded with which systems. The design is represented by the chart and the Analysis of Variance (ANOVA) table in Figure 1. Since each tape had been observed by two observers, as a check on the Medley and Mitzel procedure, a Spearman rank order correlation was computed to estimate the rater agreement between the two observers coding each tape.

			Classroom				
Taping	Recorder*	Subject	1	2	3	4	5
2	A	R					
		M					
	B	R					
		M					
3	A	R					
		M					
	B	R					
		M					
4	A	R					
		M					
	B	R					
		M					

ANOVA Table

	Source	d/f	S.S.	M.S.
1.	Classes (C)	4		
2.	Recorders (R)	1*		
3.	Subject (S)	1		
4.	Taping (T)	2		
5.	C x R	4*		
6.	C x T	8		
7.	C x S	4		
8.	R x T	2*		
9.	R x S	1*		
10.	T x S	2		
11.	C x R x T	8*		
12.	C x R x S	4*		
13.	C x S x T	8		
14.	R x T x S	2*		
15.	Residual	8*		

*Based on 2 recorders for TI; the 3 recorders for TF would have a recorder C listed and have different degrees of freedom in the ANOVA Table.

Figure 1. General reliability design for APPLE.

Interaction Analysis. Although Interaction Analysis has been tested and found to have high inter-rater reliability, there will be a small-scale attempt to study the inter-rater reliability of the project recorders. Four observers who had previous training and practice with I.A. were employed to code the tapes. The tapes of six of the 12 teachers were randomly selected from the time three taping. All observers coded both the reading and mathematics lessons. The design for use with the Medley and Mitzel technique may be represented by the chart and the ANOVA table in Figure 2.

Flanders (1967) has suggested Scott's "Pi" as an alternative method for estimating inter-rater reliability with I.A. This statistic was computed for each recorder pair coding each of the six teachers.

Rating scales. The rating scales were tested for inter-rater reliability with one of the above designs. The Videotape Rating Schedule #1 was completed by the I.A. recorders and was evaluated with the I.A. ANOVA design. Because they were completed by the APPLE recorders, the Observation-Classroom Summary Record: APPLE System and the Videotape Rating Schedule #2 were evaluated with the APPLE ANOVA design.

		Classroom					
Recorder	Subject	1	2	3	4	5	6
A	R						
	M						
B	R						
	M						
C	R						
	M						
D	R						
	M						

ANOVA Table

Source	d/f	S.S.	M.S.
1. Classes (C)	5		
2. Recorders (R)	3		
3. Situations (S)	1		
4. C x R	15		
5. C x S	5		
6. R x S	3		
7. Residual	15		

Figure 2. Reliability design for Interaction Analysis.

Results of Subsample Inter-Rater Reliability

The ANOVA design yields a statistic, Rho (ρ_{xx}), which is discussed by Medley and Mitzel. Rho is an estimate of reliability that may be interpreted similarly to traditional inter-observer correlation coefficients. However, this estimate is a more global one, taking account of all the observers over all of the classrooms recorded. The ANOVA has been calculated for each of the categories in the observation systems and each of the items on the rating scales. Tables 1-7 present the results of the analyses in a summary fashion. The first column of each table lists the Rho for each item. Rho values of .70 or higher are usually considered to be indicators of high reliability. Items with lower coefficients may also be useful, however. The second column of Table 1 and Table 2 presents the obtained Spearman's rank order correlation between two observers on all of the tapes coded. Although some of the assumptions governing the strict use of this statistic have not been met, it does allow for a rough estimate of inter-rater reliability to compare with the results of the recorder main effect.

The analysis of variance procedure which produced the Rho statistic permits, through the examination of various F ratios, the reader to speculate about <u>why</u> a particular Rho is large or small. The rest of the table contains such information about the contributions to reliability of classrooms, recorders, subject matter, and interactions. It is not enough to focus only on the reliability of raters when other factors are involved.

The next column in the tables lists the results of the F test on the classroom main effect, answering the question: are classrooms rated differently on each item? If the item or category is a sensitive one, the classrooms will be rated differently, for each classroom is unique. A behavior will be predictive of achievement only if there are differences in the observed frequency or in the observed nature of the behavior from classroom to classroom. Although we are looking only at five and six classrooms, if there were no differences in the frequency of the observed behavior there would have to be no differences in the achievement of the pupils in the five classrooms for the item to be useful. If there is a "yes" listed in this column, Rho is more likely to be large.

The next column gives the results of the F test on the recorder main effect, dealing with whether recorders are in agreement. They should be in agreement, of course, if the item has high reliability and if the recorders are recording with equal frequency. The Rho statistic may be small, even though there is a "yes" listed in this column. Although the observers are in agreement, other situational factors may cause an item or category to be unreliable. These factors are reflected in the questions and answers listed in the other columns.

The next column listed indicates if reading and mathematics are rated differently on this item as estimated by the results of the F test on the main effect of the subject. A "yes" or a "no" in the fourth column is of interest only if one is concerned with

the item discriminating between a reading and a mathematics lesson. On the other hand, if reading and mathematics lessons are rated differently, the Rho statistic will be deflated.

The next column answers the question: do different observers rate different classes similarly? This question is answered by an examination of the interaction C x R. If recorders are consistent, they should rate them the same.

The next column lists an answer to: do different observers see reading and mathematics lessons similarly? This question is answered by an examination of the interaction R x S. If there is inter-rater reliability, they should see essentially the same events in the same classrooms.

Where time was included in the design, the next to last column presents the results of the test of the main effect for time. The question answered is: are the events present in the same frequency or are the ratings the same for the classrooms across time? A "yes" in this column indicates that what is being observed is stable across time, or at least, that the observers are consistent across time. (This question is addressed in a different design in the next section.)

Tables 1 and 2 contain a final column with the mean number of observations of each category per tape over all the tapes, not just those for the subsample. This number is also a guide to the frequency with which a particular event occurs in a lesson. Those categories with low frequencies may not be accurately evaluated by the Medley and Mitzel technique, or by other common procedures of

estimating reliability. In the tables, those items with high Rho values and with "yes" in the rest of the columns may be considered the most reliable items.

APPLE. Table 1 presents the results of the calculations on the Teacher Instructional Lexicon events. In general, Rho is quite low for most of the events but the statistic may be misleading because of the rarely observed events. Events such as 3, 4E, 4F, 4G, 5C, 5D, 5E and 5F were noted so infrequently that often the statistics could not be calculated. For these events, the Spearman's r statistic serves to indicate which of them are more reliable than the others. Although exceedingly rare, 3, 4F, and 5D, seem to have high inter-rater reliability. Event 4A1 was common but because of an unusual circumstance Rho could also not be obtained. The Spearman statistic does indicate reliability in this case, however. All told, only events 4D2, 4H, 5A, and 5G have Rho values greater than .50.

Column three of the table indicates there is no variation among the classrooms for approximately half the events. If there is no variation indicated among these five teachers, the event might not be a good one for predicting differences in achievement.

The large number of "no" entries in column four indicates that for many events one of the observers noted the occurrence more frequently than the other. On items 4D2 and 5A where there are large Rho's it is probable that one observer was quite consistently different from the other in recording more of each event on each tape. Except for event 4F, 4G, 5BD, 5D and 5E, there is reasonable agreement between Rho and Spearman's r.

Columns five and six show only a few significant interactions. Most of the items were observed reliably in these two respects.

The penultimate column indicates by the number of "yes" entries that most of the events occurred to about the same extent on each of the tapings. While one would expect variation in lessons from time to time, it is important that variables abstracted from one-time videotaping (i.e. as in the main sample) be variables that are representative, in general, of what happens in lessons thoughout the year. Some of the variables in this lexicon do not seem to be stable aspects of the teacher's behavior. These differences across time contribute to the low Rho values in many of the events. The next major section on consistency examines the events across time with another ANOVA design as verification of these tentative findings.

Looking at the table as a whole, few of these instructional event categories seem to hold much promise in their present form. It may be possible to increase the reliability of the system by combining some of the present categories. Some of the distinctions between events could have been unclear to the observers. It may also be true that the events are simply so hard for recorders to observe on videotape that these aspects of teacher behavior are best defined by the teacher or pupil and not by an external observer.

From Table 1 it appears that the events which seem most useful are 4H (Direct Cueing) and 5G (Asking if Information Understood), although 4A1 (Stating Facts on the Topic), 4D1 (Giving Directions on the Topic), 5A (Asking for Answer), 5BP (Asking for the Development of a Response) and 5C (Asking for Agreement) could also prove to have some utility.

Table 1

Inter-rater Reliability of Two Recorders Coding Five Teachers During Reading and Mathematics Lessons on Three Occasions Using Teacher Instruction Lexicon

		General Estimate of Reliability in Inter-rater Situation (Rho, ρ_{xx})	Spearman's r between observers	Are classrooms different on this item? (yes = F_c sig. $p \leq .05$)	Do recorders rating this item agree? (yes = F_r not sig. $p \leq .05$)	Do mathematics lessons and reading lessons have the same number of these events (yes = F_s not sig. $p \leq .05$)	Do different observers rate different classes similarly? (yes = F_{cxr} not sig. $p \leq .05$; or σ^2_{cr} is small)	Do different observers see reading and mathematics lessons the same? (yes = F_{rxs} not sig. $p \leq .05$; or σ^2_{rs} is small)	Are the tapes rated similarly across time (yes = F_t not sig. $p \leq .05$)	Mean observed event frequency per tape (Total Sample N = 186)
1.	Gaining Attention	.26	.44	Yes	No	Yes	Yes	Yes	Yes	.62
2.	Goal Statement	.40	.21	No	No	Yes	Yes	Yes	Yes	.41
3.	Stimulating Recall	.00	.74	No	Yes	Yes	Yes	Yes	Yes	.30
4.A1	Stating Fact on Topic	.00*	.76	Yes	Yes	No	No	No	No	14.44
4.A2	Stating Fact off Topic	.07	.46	No	No	Yes	Yes	Yes	Yes	.60
4.B	Giving an Example	.00	.07	No	No	Yes	No	No	Yes	.56
4.C	Giving an Explanation	.32	.15	Yes	No	Yes	No	No	No	1.04
4.D1	Giving Directions on Topic	.20	.52	Yes	Yes	No	No	Yes	Yes	8.22
4.D2	Giving Directions off Topic	.54	.25	No	No	Yes	Yes	Yes	Yes	1.81
4.E	Summarizing	.00	.29	No	No	Yes	Yes	Yes	Yes	.81
4.F	Giving Application	.00	1.00	No	No	Yes	Yes	Yes	Yes	.25

*The calculation here = 0 because of a large classroom x time interaction

Table 1 (continued)

Inter-rater Reliability of Two Recorders Coding Five Teachers During Reading and Mathematics Lessons on Three Occasions Using Teacher Instruction Lexicon

		General Estimate of Reliability in Inter-rater Situation (Rho, ρ_{xx})	Spearman's r between observers	Are classrooms different on this item? (yes = F_c sig. $p \leq .05$)	Do recorders rating this item agree? (yes = F_r not sig. $p \leq .05$)	Do mathematics lessons and reading lessons have the same number of these events (yes = F_s not sig. $p \leq .05$)	Do different observers rate different classes similarly? (yes = F_{cxr} not sig. $p \leq .05$; or σ^2_{cr} is small)	Do different observers see reading and mathematics lessons the same? (yes = F_{rxs} not sig. $p \leq .05$; or σ^2_{rs} is small)	Are the tapes rated similarly across time (yes = F_t not sig. $p \leq .05$)	Mean observed event frequency per tape (Total Sample N = 186)
4.G	Contrasting and Comparing	.32	.26	No	Yes	Yes	Yes	Yes	Yes	.29
4.H	Direct Cueing	.59	.66	Yes	Yes	Yes	Yes	No	No	1.06
5.A	Asking for Answer	.85	.84	Yes	No	Yes	Yes	Yes	No	23.82
5.BA	Asking for Additional Information	.34	.55	Yes	No	Yes	Yes	Yes	No	4.94
5.BD	Asking for Development of a Response	.35	.70	Yes	No	No	No	Yes	Yes	2.76
5.BP	Asking for Application of a Response	.24	.40	Yes	Yes	Yes	Yes	Yes	Yes	.38
5.C	Asking for Agreement	.31	.40	Yes	Yes	No	Yes	Yes	Yes	.77
5.D	Asking for Disagreement	.00	1.00	No	No	Yes	Yes	Yes	Yes	.13
5.E	Asking for a Judgement	.21	-.11	No	Yes	Yes	Yes	Yes	Yes	.51
5.F	Asking for Performance Appraisal, Testing	.00	.38	No	Yes	Yes	Yes	Yes	Yes	.15
5.G	Asking if Information Understood	.57	.68	Yes	Yes	No	Yes	Yes	No	1.34

Table 2 contains the results of the calculations on the Teacher Feedback events. Again there are few events with Rho values greater than .50 (1A, 1B, 1C, 1D3, 2B3, 3A1, 3C, 3D6, 3F and 3H1). Spearman's r, however, indicates a few more of the low frequency events which may have inter-rater reliability (1D4, 1E, 2F, 3D8, 3H2). Many events (see third column) do differ in frequency across classrooms. But of the tentatively reliable events, events 1E, 2F, 3D8 and 3H2, do not have significant F values for the classroom main effect, probably because of their low rate of occurance. Column four indicates that there were a number of events where at least one of the observers was recording more events of a kind than the others. In many instances, however, this fact did not seem to influence the correlation between observers (see column two). The fifth column in Table 2 indicates that there are differences between reading and mathematics lessons for 8 of 35 events. The next to last column indicates that about half of the events are not reliable across time, including events 1B, 1C, 2B3, 3A1, and 3D6. The next section checks this finding by means of a different design.

In their present form, using a strict criteria (high Rho and r values and "yes" across the columns), only one event, 1D3 (Positive Physical Contact) seems to be highly reliable. Other events, 1A, 2B3, 3D6, and 3F, may also prove useful.

Table 2

Inter-rater Reliability of Three Recorders Coding Five Teachers During Reading and Mathematics Lessons on Three Occasions Using Teacher Feedback Lexicon

Teacher Feedback Event	General estimate of reliability in inter-rater situation. (Rho, ρ_{xx})	Spearman's r between observers*	Are classrooms (teachers) different on this item? (yes = F_c sig. $p \leq .05$)	Are observers recording the same number of events? (yes = F_r not sig. $p \leq .05$)	Do mathematics and reading lessons have the same observed frequency of this event? (yes = F_s not sig. $p \leq .05$)	Do different observers code different classes the same? (yes = $F_{c \times r}$ not sig. $p \leq .05$; or δ^2_{cr} is small)	Do different observers code reading and mathematics lessons the same? (yes = $F_{r \times s}$ not sig. $p \leq .05$; or δ^2_{rs} is small)	Are the tapes rated similarly across time? (yes = F_t not sig. $p \leq .05$)	Mean observed frequency per tape (Total Sample, N=186)
1. A Simple knowledge of results	.85	.87	yes	no	no	no	yes	yes	67.35
1. B Elaborated knowledge of results	.56	.47	yes	no	no	no	yes	no	8.14
1. C Directions to child to elaborate	.52	.53	yes	no	yes	no	yes	yes	5.72
1. D1 General verbal praise	.31	.38	yes	no	yes	no	no	no	10.54
1. D2 Granting privileges	.31	.28	yes	no	yes	no	yes	no	.02
1. D3 Positive physical contact	.82	.70	yes	yes	yes	yes	yes	yes	1.06
1. D4 Award of consumables, tokens or symbols of recognition	.30	.54	yes	yes	no	yes	no	no	.73
1. E Teacher directs others to give feedback	.00	1.00	no	yes	yes	yes	yes	yes	.26
1. F Ignoring of positive behavior or achievement	.00	.09	yes	yes	yes	yes	yes	no	1.45
2. A Acknowledges hearing response	.00	.21	no	no	yes	no	yes	yes	21.24

Table 2 (continued)

Inter-rater Reliability of Three Recorders Coding Five Teachers During Reading and Mathematics Lessons on Three Occasions Using Teacher Feedback Lexicon

Teacher Feedback Event	General estimate of reliability in inter-rater situation. (Rho, ρ_{xx})	Spearman's $\underline{r}$ between observers*	Are classrooms (teachers) different on this item? (yes = F_c sig. $p \leq .05$)	Are observers recording the same number of events? (yes = F_r not sig. $p \leq .05$)	Do mathematics and reading lessons have the same observed frequency of this event? (yes = F_s not sig. $p \leq .05$)	Do different observers code different classes the same? (yes = $F_{c \times r}$ not sig. $p \leq .05$; or $\hat{\sigma}^2_{cr}$ is small)	Do different observers code reading and mathematics lessons the same? (yes = $F_{r \times s}$ not sig. $p \leq .05$; or $\hat{\sigma}^2_{rs}$ is small)	Are the tapes rated similarly across time? (yes = F_t not sig. $p \leq .05$)	Mean observed frequency per tape (Total Sample, N=186)
2. B1 Indication of both positive and negative elements in response both weighted equally	.00	-.02	no	no	yes	yes	yes	yes	6.81
2. B2 Both, but emphasis on the positive	.09	-,08	no	no	yes	yes	yes	yes	12.95
2. B3 Both, but emphasis on the negative	.54	.26	yes	no	no	yes	yes	no	26.72
2. C Direction to child to elaborate	.00	.11	yes	no	yes	no	yes	no	9.68
2. D Promise of reward contingent on change or continuation	.00	.31	yes	no	yes	yes	yes	no	.20
2. E Teacher directs others to give feedback	.37	.41	yes	yes	yes	yes	yes	yes	1.87
2. F Ignores response	.17	.62	no	yes	yes	yes	yes	yes	2.14
3. A1 Answer termed wrong	.52	.61	yes	yes	no	no	no	no	6.75

Table 2 (continued)

Inter-rater Reliability of Three Recorders Coding Five Teachers During Reading and Mathematics Lessons on Three Occasions Using Teacher Feedback Lexicon

Teacher Feedback Event	General estimate of reliability in inter-rater situation. (Rho, ρ_{xx})	Spearman's r_s between observers*	Are classrooms (teachers) different on this item? (yes = F_c sig. $p \leq .05$)	Are observers recording the same number of events? (yes = F_r not sig. $p \leq .05$)	Do mathematics and reading lessons have the same observed frequency of this event? (yes = F_s not sig. $p \leq .05$)	Do different observers code different classes the same? (yes = F_{cr} not sig. $p \leq .05$; or σ^2_{cr} is small)	Do different observers code reading and mathematics lessons the same? (yes = F_{rs} not sig. $p \leq .05$; or σ^2_{rs} is small)	Are the tapes rated similarly across time? (yes = F_t not sig. $p \leq .05$)	Mean observed frequency per tape (Total Sample, N=186)
3. A2 Correct answer given	.08	.15	yes	no	yes	no	yes	yes	6.84
3. B1 Error elaborated	.31	.17	yes	no	yes	yes	yes	yes	2.20
3. B2 Correct answer elaborated	.45	.12	yes	no	no	yes	no	no	.77
3. C Direction to try again	.52	.60	yes	no	yes	yes	yes	yes	7.76
3. D1 Verbal reprimand	.36	.49	yes	no	yes	no	yes	no	6.23
3. D2 Withdrawal of privileges	.00	-.10	no	yes	yes	yes	no	yes	.17
3. D3 Aversive physical contact	.00	-.02	yes	yes	yes	yes	yes	no	.23
3. D4 Denial of consumables, tokens or symbols of recognition	.00	-	-	-	-	-	-	-	.01
3. D5 Directions to stop activity	.00	.08	no	no	yes	no	yes	yes	2.14
3. D6 Warning of consequences	.88	.73	yes	yes	yes	yes	no	no	.73
3. D7 Directions to proceed with assigned activity	.00	.46	yes	no	no	yes	no	no	2.74

Table 2 (continued)

Inter-rater Reliability of Three Recorders Coding Five Teachers During Reading and Mathematics Lessons on Three Occasions Using Teacher Feedback Lexicon

Teacher Feedback Event	General estimate of reliability in inter-rater situation. (Rho, ρ_{xx})	Spearman's r between observers*	Are classrooms (teachers) different on this item? (yes = F_c sig. $p \leq .05$)	Are observers recording the same number of events? (yes = F_r not sig. $p \leq .05$)	Do mathematics and reading lessons have the same observed frequency of this event? (yes = F_s not sig. $p \leq .05$)	Do different observers code different classes the same? (yes = $F_{c \times r}$ not sig. $p \leq .05$; or σ^2_{cr} is small)	Do different observers code reading and mathematics lessons the same? (yes = $F_{r \times s}$ not sig. $p \leq .05$; or σ^2_{rs} is small)	Are the tapes rated similarly across time? (yes = F_t not sig. $p \leq .05$)	Mean observed frequency per tape (Total Sample, N=186)
3. D8 Isolation, time out, sent from room	.00	.79	no	yes	yes	yes	yes	no	.53
3. E Teacher directs others to give feedback	.11	.41	no	no	yes	yes	yes	yes	.82
3. F Ignoring of negative behavior	.75	.57	yes	no	yes	no	yes	yes	.36
3. G Teacher acknowledges defeat	.00	1.00	-	-	-	-	-	-	.00
3. H1 Acceptance of pupil's lack of answers	.50	.54	yes	no	no	yes	yes	yes	1.77
3. H2 Rejection	.35	1.00	no	yes	yes	yes	yes	yes	.34

* between the two observers who coded the major sample

Interaction Analysis. Table 3 contains the inter-rater reliability results for Interaction Analysis. Interaction Analysis was included in the present study to provide a benchmark against which the other observation systems might be compared. Inspection of the table indicates that the recordings of the ten I.A. categories were not equally reliable. Only two could be considered to have high reliability: 4-"Asks Questions" and 10-"Silence or Confusion". Although generally reliable, 6-"Gives Directions" was not always recorded the same way in the same classrooms by different observers. Category 2-"Praises or Encourages", 3-"Accepts or Uses Ideas of Pupils" and 9-"Pupil Talk-Initaition" were not reliable. Category 1-"Accepts Feelings" and Category 8-"Pupil Talk-Response" have lower reliability but may be useful. Both 5-"Lecturing" and 3-"Accepts or Uses Ideas of Pupils" were rated with unequal frequency by the observers; one or more of the observers was more apt to observe this on the tapes than others. Table 4 presents the average inter-rater reliability figures calculated with Scott's "Pi". The Pi's are nowhere near the desirable .70 point and validate the findings represented in Table 3.

As the Interaction Analysis coders had had experience in other live and recorded settings where they had demonstrated their ability to code reliably, we are left to speculate about what it is in the present videotape medium which caused them difficulty. Because the microphone was attached to the teacher, it is not too surprising that they had trouble determining "Pupil Talk-Initiation" because they probably could not hear it very well. Other than the problem

in hearing the children and the fact the content of the tapes was very natural with no special planned lesson, many times perhaps there were situations present on the tape that are not found on the usual interaction analysis study. In addition, the lessons coded were quite long, up to 40 minutes. Most uses of I.A. do not involve this lengthy a coding session.

Rating scales. Table 5 presents the inter-rater reliability results for Videotape Rating Schedule #1. Three of the items meet the criteria of high Rho values and a "yes" across the columns and may be considered to be reliable: Teacher Movement, Teacher Gestures, and Teacher Enthusiasm. Although they were not rated as reliably, Time Spent Learning Activities, Lesson Learning Centered, and Differentiation of Learning Activities may also yield useful results.

Table 6 contains the calculations for the five items on the Videotape Rating Schedule #2. The rating of the attentiveness of the children was the most reliable item followed by the judgement of whether or not the teacher routinely named the child who was to answer as the teacher asked a question. In general the rating of teacher warmth was quite consistent, but evidently the observers did not agree at all about one teacher of the five. It would probably be best to consider this item unreliable for future use.

The inter-rater reliability data for the Observation Summary Record: APPLE System is presented in Table 7. This table does not contain the data found in columns four and six of the other tables since the reading and mathematics tapes were not rated independently with this instrument. Several of the items on this summary schedule seemed to have good inter-rater reliability: Frequent Praise,

Praise to Many, Adults Initiate Interaction, Students in Control, Need for Intervention, Students Require Motivation, Adults Negative. One other item had lower reliability but may be a useful measure: Students Initiate Interaction. Two-thirds of the items on this questionnaire did not prove to have good inter-rater reliability.

In summary, the rating scales seem to be the most reliable measures of the videotapes. It must be remembered, however, that the rating scales only have four choices and that this fact reduces the range of discrepancies that the observers may show. Nevertheless, observation systems applied to videotapes of the particular classrooms in this study (typical second and fifth grade classrooms) with these recorders (graduate students in education) have not yet been developed to the point where they are very reliable.

Table 3

Inter-rater Reliability of Four Recorders Coding
Six Teachers on Time Three Tapes
with Interaction Analysis

Category Label	General estimate of reliability in inter-rater situation (ρ_{xx})	Are classrooms (teachers) different on this item? (Yes-F_c sig. $p \leq .05$)	Do recorders rating this item agree? (Yes-F_r not sig. $p \geq .05$)	Do mathematics classes and reading lessons have the same number of these events? (Yes-F_s not sig. $p \geq .05$)	Do different observers rate different classes the same? (Yes-F_c x r not sig. $p \geq .05$ or, $\delta^2_{c \times r}$ is small)	Do different observers see reading and mathematics lessons the same? (Yes-F_r x s not sig. $p \geq .05$ or, $\delta^2_{r \times s}$ is small)
1. Accepts feelings	.660	no	yes	yes	yes	yes
2. Praises or encourages	.324	no	no	no	yes	yes
3. Accepts or uses ideas of pupils	.479	yes	no	no	yes	no
4. Asks questions	.691	yes	yes	no	yes	yes
5. Lecturing	.551	yes	no	yes	yes	yes
6. Gives directions	.714	yes	no	no	no	yes
7. Criticizing or justifying authority	.635	yes	no	no	yes	yes
8. Pupil talk-response	.623	yes	yes	yes	yes	yes
9. Pupil talk-initiation	.044	no	no	yes	yes	yes
10. Silence or confusion	.883	yes	yes	no	yes	yes

Table 4

Average Inter-Rater Reliabilities Across Six Teachers with Interaction Analysis Calculated by Scott's Pi

Reading					Mathematics				
Recorder	P	D	B	R	Recorder	P	D	B	R
D	.466	--	--	--	D	.543	--	--	--
B	.592	.500	--	--	B	.408	.469	--	--
R	.656	.541	.626	--	R	.492	.658	.373	--

Table 5

Inter-Rater Reliability of Four Recorders Rating Six Teachers on Time Three Tapes with Videotape Rating Schedule #1

Item	Content	General estimate of reliability in inter-rater situation ($\hat{\rho}_{xx}$)	Are classrooms (teachers) different on this item? (Yes=F_C sig. $p \leq .05$)	Do recorders rating this item agree? (Yes=F_r not sig. $p \leq .05$)	Do mathematics lessons and reading lessons have the same number of these events? (Yes=F_s not sig. $p \leq .05$)	Do different observers rate different classes the same? (Yes=F_{Cxr} not sig. $p \leq .05$ or, σ^2_{2cxr} is small)	Do different observers see reading and mathematics lessons the same? (Yes=F_{rxs} not sig. $p \leq .05$ or, σ^2_{rxs} is small)
1.	Variability During Lesson	.32	yes	yes	no	yes	yes
2.	Pace of Lesson	.18	no	no	no	yes	yes
3.	Tempo of Lesson		Discarded by recorders as too difficult to rate				
4.	Teacher Movement	.73	yes	yes	no	yes	yes
5.	Teacher Gestures	.68	yes	yes	no	yes	yes
6.	Pupil Enjoyment	.83	yes	no	no	no	yes
7.	Teacher Enthusiasm	.90	yes	yes	no	yes	yes
8.	Structuring of Learning Environment	.23	yes	yes	yes	no	yes
9.	Time Spent on Learning Activities	.53	yes	yes	no	yes	yes
10.	Teacher Organization	.77	yes	no	yes	yes	yes
11.	Lesson Learning-Centered	.57	yes	yes	yes	yes	yes
12.	Differentiation of Learning Activities	.56	yes	yes	yes	yes	yes
13.	Teacher Sensitivity to Feelings		Discarded by recorders as too difficult to rate				
14.	Body Language Indicating Emotional Support		Discarded by recorders as too difficult to rate				

Table 6

Inter-rater Reliability of Three Recorders Rating Five Teachers on Three Tapings of Reading and Mathematics with Videotape Rating Schedule #2

	General Estimate of Reliability in Inter-rater Situation (ρ_{xx})	Are classrooms different on this item? (yes = F_C sig. $p \leq .05$)	Do recorders rating this item agree? (yes = F_r not sig. $p \leq .05$)	Do mathematics lessons and reading lessons have the same number of these events? (yes = F_S not sig. $p \leq .05$)	Do different observers rate different classes similarly? (yes = F_{cxr} not sig. $p \leq .05$; or $^2_{cr}$ is small)	Do different observers see reading and mathematics lessons the same? (yes = F_{rxs} not sig. $p \leq .05$; or $^2_{rs}$ is small)
1. How attentive are children?	.8704	yes	yes	yes	yes	yes
2. Does teacher name child to answer when asking questions?	.5719	yes	yes	yes	yes	yes
3. How long does teacher wait for reply?	.2908	yes	yes	no	no	yes
4. Teacher warmth	.9398	yes	no	yes	no	yes
5. Teacher honesty	.6304	yes	no	yes	no	yes

Table 7

Inter-rater Reliability of Two Recorders Rating Five Teachers on Three Tapings of Mathematics and Reading Lessons Using Observation Classroom Summary Record: APPLE System

Item	Content	General estimate of reliability in inter-rater situation (ρ_{xx})	Are classrooms (teachers) different on this item? (yes = F_c sig. $p \leq .05$)	Do recorders rating this item agree overall? (yes = F_r not sig. $p \leq .05$)	Do different observers rate different classes similarly? (yes = F_{cxr} not sig. $p \leq .05$ or ρ^2_{cxr} is small)	Are the tapes rated similarly across time? (yes = F_r not sig. $p \leq .05$)
1.	Frequent Praise	.948	yes	yes	yes	yes
2.	Praise to many	.925	yes	yes	yes	yes
3.	Students initiate interactions	.641	yes	yes	yes	yes
4.	Adults initiate interaction	.815	yes	yes	yes	yes
5.	Students respond	.936	yes	no	yes	yes
6.	Students correct	.195	no	no	yes	yes
7.	Corrective feedback	.814	yes	no	no	yes
8.	Disregard observer	.851	yes	no	yes	yes
9.	Adults organize groups	.548	no	yes	yes	yes
10.	Previously organized groups	.621	no	yes	yes	yes
11.	Same Subject	.465	no	yes	yes	yes
12.	Students decide	.322	no	yes	yes	yes
13.	Students in control	.819	yes	yes	yes	yes
14.	Need for intervention	.934	yes	yes	yes	no
15.	Adults instruct	.091	yes	yes	yes	yes
16.	Students independent	.000	no	no	no	yes
17.	Students in large groups	.540	no	no	yes	yes
18.	Students in small groups	.254	no	no	yes	yes
19.	Students self directed	.920	yes	no	yes	yes
20.	Students require motivation	.732	yes	yes	yes	yes
21.	Students friendly	.353	no	no	yes	yes
22.	Occasionally unfriendly	.376	yes	yes	yes	no
23.	Pupil progress monitored	.699	yes	no	yes	yes
24.	Adults negative	.970	yes	yes	yes	yes

The Consistency of Subsample Teacher Behavior Over Time

Procedures. To judge whether or not the teacher behaviors on videotapes are consistent aspects of the teacher's style over time, the teachers were videorecorded on four occasions. If the data collected on each of these occasions are similar, then videotaping does have promise as a technique for gathering certain kinds of data on teaching.

Not all of the tapes were coded with all of the observation systems or rating schedules. APPLE observers coded the tapes of times two, three and four. Interaction Analysis was only used with time two and time three tapes. Some of the time one tapes were coded; but the quality of the tapes was so poor that these data were not used in the analysis. Videotape Rating Schedule #1 was completed on the first three tapings and the other rating scales on the time two, three and four tapes.

The consistency of behaviors across time was evaluated with an ANOVA nested design. The nested design was selected because there was not enough time to have more than one observer code all of the tapes. The general design is represented in Figure 3. In particular evaluations the number of recorders varied from 2 to 4, the number of teachers from 10 to 12, and the number of tapings from 1 to 3. "Subject" was missing from one design where both reading and mathematics were rated together.

Results of Consistency of Subsample Teacher Behavior Over Time

The F tests computed from this analysis yielded a wealth of information. We were primarily interested in whether teachers are consistent over time. To learn whether this was so we computed the F ratio for the main effect of taping and examined the interactions involving taping (T). The summary results of the analysis for each reliable item are presented in Tables 8-12. It was not necessary to evaluate all of the categories or items. Only the items that had demonstrated inter-rater reliability were used to evaluate classroom consistency.

The first question addressed in each table is, within each classroom, is this item rated the same on each of the tapings? An affirmative answer is listed in the table where the F ratio for the main effect of taping is not significant. A "yes" answer indicates that the item is consistent over time ($p \leq .05$).

The second question addressed in the tables is, do reading and mathematics lessons differ in consistency over time on this item? This question is answered by $F_{T \times S}$. A "no" response indicates that they do not differ. The third question addressed is, do recorders differ in how consistent they are over time? It is answered by an examination of the F ratio for the T x R interaction.

APPLE. From Table 8 where the results for the consistency study with APPLE events are displayed, it appears that teachers are consistent on eight of the behaviors. The last two behaviors, 2B2 and 3F, plus 5C, (which has doubtful reliability) have "yes" in the second column indicating that the recorders differ in how consistently

they perceive these behaviors in their classrooms over time. These three behaviors should not be considered capable of being consistently observed. In the other design, with fewer teachers, events 4H, 5G, 4A1, 5A and 3D6, were not consistent across time whereas in this design they are. There is evidence that these events are stable aspects of teacher behavior.

Interaction Analysis. Table 9 displays the summary of the consistency analysis of Interaction Analysis. All of the categories were included in the table, even the ones with low inter-rater reliability. Within raters, however, an inspection of the table indicates that all of the categories represent behaviors which are consistent over time.

Rating scales. Table 10 summarizes the consistency study of Videotape Rating Schedule #1. From the table it is clear there is no reason to suspect that the teachers are inconsistent in the aspects of their teaching style tapped by these items. They are consistent in both their reading and mathematics lessons in the aspects listed. On one item, 11 (Lesson Learning-Centered), one or more of the recorders was more consistent than his or her colleagues but as the item has high inter-rater reliability, this fact should not trouble anyone.

Table 11 contains the summary of the consistency study of all of the items on Videotape Rating Schedule #2. The first item, How Attentive Are the Children, has higher inter-rater reliability yet raters differ in the extent to which they see it as consistent over time although as the "yes" in column one indicates, they do see this behavior as consistent. Interestingly enough, Teacher Warmth does

not seem to be a consistent aspect of the teachers on the tapes. It makes sense that they would be more accepting and open to children on some days rather than on others, for we all have swings in mood. The other items on Videotape Rating Schedule #2 seem to be getting at consistent teacher behaviors.

Table 12 presents the results of the consistency study of the reliable items of the Observation Summary Record: APPLE System. The table does not contain a column three as there were no separate ratings of reading and mathematics over time. According to this table, all of these items are measures of behaviors that are consistent over time.

	Recorder	A						B					
	Classroom	1	2	3	4	5	(6)	7	8	9	10	11	(12)
	Subject												
Taping 2	Reading												
	Math												
Taping 3	Reading												
	Math												
Taping 4	Reading												
	Math												

ANOVA Table

Source	d/f*	S.S.	M.S.	F.
Between Classrooms	9			
1. Between Recorders	1			
2. Residual I	8			
Within Classrooms	50			
1. Subject (S)	1			
2. Taping (T)	2			
3. T x S	2			
4. T x R	2			
5. S x R	1			
6. T x S x R	2			
7. Residual II	40			
Total	59			

*The degrees of freedom vary with the number of levels of the various factors in the design such as number of recorders or number of teacher classrooms.

Figure 3. General design for studying consistency.

Table 8

Consistency Over Three Tapings of Ten Teachers on Eleven Events From APPLE Teacher Instruction Lexicon and Teacher Feedback Lexicon

Item	Within each classroom is this event observed with equal frequency on each taping? (Yes = F_T not sig. $p \leq .05$)	Do recorders differ in how consistent they are in observing this event? (Yes = $F_T \times R$ sig. $p \leq .05$)	Do reading and mathematics lessons differ in consistency over time on this event? (Yes = $F_T \times S$ sig. $p \leq .05$)
T.I.			
4H Direct cueing	yes	no	no
5G Asking if information understood	yes	no	no
*4AI Stating facts on the topic	yes	no	no
*5A Asking for answer	yes	no	no
*5BP Asking for the development of a response	yes	no	no
*5C Asking for agreement	yes	yes	yes
T.F.			
1D3 Positive physical contact	yes	no	yes
3D6 Warning of consequences	yes	no	no
*1A Simple knowledge of results	yes	no	no
*2B2 Indication of both positive and negative elements in the response but emphasis on negative	yes	yes	no
*3F Ignoring of negative behavior	no	yes	yes

*Less reliable items

Table 9

Consistency Over Two Tapings of 12
Teachers on 10 Categories of Interaction Analyis

Category Label	Within each classroom, is this category observed with equal frequency on each taping? (yes = F_T not sig. $p \leq .05$)	Do recorders differ in how consistent they are on this item? (yes = $F_{T \times R}$ sig. $p \leq .05$)	Do reading and mathematics differ in consistency over time on this item? (yes = $F_{T \times S}$ sig. $p \leq .05$)
1. Accepts Feelings	yes	no	no
2. Praises or Encourages	yes	no	no
3. Accepts or Uses Ideas of Pupils	yes	no	no
4. Asks Questions	yes	no	no
5. Lecturing	yes	no	no
6. Gives Directions	yes	no	no
7. Criticizing or Justifying Authority	yes	no	no
8. Pupil Talk-Response	yes	no	no
9. Pupil Talk-Initiation	yes	no	no
10. Silence or Confusion	yes	no	no

Table 10

Consistency Over Three Tapings of 12 Teachers on Six Items from Videotape Rating Schedule #1 with High Inter-Rater Reliability

Item	Content	Within each classroom, is this item rated the same on each taping? (Yes = F_T not sig. $p \leq .05$)	Do recordings differ in how consistent they are over time on this item? (Yes = F_{TxR} sig. $p \leq .05$)	Do reading and mathematics differ in consistency over time on this item? (Yes = F_{TxS} sig. $p \leq .05$)
4.	Teacher Movement	Yes	No	No
5.	Teacher Gestures	Yes	No	No
7.	Teacher Enthusiasm	Yes	No	No
9.	Time Spent on Learning Activities	Yes	No	No
11.	Lesson Learning-Centered	Yes	Yes	No
12.	Differentiation of Learning Activities	Yes	No	No

Table 11

Consistency Over Three Tapings of 10 Teachers on Videotape Rating Schedule #2

Item Content	Within each classroom, is this item rated the same on each taping? (yes = F_T not sig. $p \leq .05$)	Do recorders differ in how consistent they are over time on this item? (yes = $F_{T \times R}$ sig. $p \leq .05$)	Do Reading and Mathematics differ in consistency over time on this item? (yes = $F_{T \times S}$ sig. $p \leq .05$)
1. How attentive are children?	yes	yes	no
2. Does teacher name child to answer when asking questions?	yes	no	no
3. How long does teacher wait for reply?	yes	no	no
4. Teacher warmth	no	yes	no
5. Teacher honesty	yes	no	no

Table 12

Consistency Over Three Tapings of 10 Teachers on Selected Items of Observation Summary Record: APPLE System

Item	Within each classroom is this item rated the same on each taping? (yes = F_T not sig. $P \geq .05$)	Do recorders differ in how consistent they are over time on this item? (yes = $F_T \times R$ sig. $P \geq .05$)
1. Frequent Praise	yes	no
2. Praise to many	yes	no
3. Students initiate interaction	yes	no
4. Adults initiate interaction	yes	no
13. Students in control	yes	no
14. Need for intervention	yes	no
20. Students require motivation	yes	no
24. Adults negative	yes	no

Frequency of Occurrence of Events in BTES Classrooms

It is valuable in and of itself to examine the frequency of occurrence of various events in the project classrooms. For years, teacher educators and educational psychologists have been promoting various teacher behaviors as good practice or as consistent with learning theory. We may well ask how successful these efforts have been in influencing what teachers actually do in the classroom.

Teacher Instruction Lexicon. How common are various instructional behaviors? Tables 13 and 14 list mean numbers of Teacher Instructional events observed on tapes of all teachers and on tapes grouped according to lesson content, grade level and socioeconomic status of the schools. From an examination of the first column labeled "Total", it is evident that teachers at the elementary level spend most of their time giving facts, asking simple questions and giving directions about how to do something. On the other hand, the tables indicate that teachers rarely stimulate recall of prerequisite skills in the process of giving a lesson, rarely give applications of ideas they are teaching, rarely contrast and compare ideas, almost never ask for negative opinions, and do not often attempt to determine if a child has mastered a concept. Yet theoretically, these are very important behaviors teachers can use to promote higher order learning and critical thinking in their pupils. The reader is left to ponder the implications of other low frequency events.

Table 13 presents data enabling us to compare frequencies of instructional events in reading lessons and mathematics lessons and

to compare second grade tapes with fifth grade tapes. Because of the large number of events, no statistical tests were made of the differences, but for the purpose of generating hypotheses for further study we may look for differences between these groups greater than approximately one-half of a standard deviation.

Reading and mathematics lessons appear to differ in that there are more explanations given in mathematics lessons whereas in reading there is more cueing and asking follow-up questions for additional information. Reading lessons would probably involve more two-way interactions between pupils and teachers than would mathematics lessons.

There does not seem to be much difference in the frequency of the various instructional events in the second and fifth grades. The substance, the content of the events, undoubtedly does differ between grade levels, but not the absolute occurrence.

Table 14 permits us to explore differences between tapes recorded in classrooms containing children from different socio-economic status groups (SES has been defined in another section of the report, Volume II). Teachers in lower class schools seem to differ from other teachers in that they are less prone to state facts, are less likely to give explanations, and are less likely to ask if information is understood. In general there seems to be somewhat less verbal behavior in these classrooms than in others. Teachers in upper middle class schools, on the other hand, state facts more often, ask more frequently if information is understood, and engage in less direct cueing than do other teachers.

Table 13

Mean Observed Frequencies of Teacher Instruction Categories:
Total Group, Reading/Math Subgroups, Grade Level Subgroups

Category Number	Category Name	Total N=187 Mean	Reading N=93 M.	Reading N=93 S.D.	Math N=93 M.	Math N=93 S.D.	Second Grade N=82 M.	Second Grade N=82 S.D.	Fifth Grade N=104 M.	Fifth Grade N=104 S.D.
1	Gaining Attention	.62	.72	.90	.53	.58	.57	.75	.66	.77
2	Goal Statement	.41	.40	.53	.42	.52	.38	.54	.43	.52
3	Stimulating Recall	.30	.24	.50	.37	.57	.30	.51	.30	.56
4A1	Stating Fact on Topic	14.44	13.94	6.48	15.10	7.56	12.68	5.36	15.96	7.86
4A2	Stating Fact off Topic	.60	.64	1.22	.56	1.02	.48	.84	.70	1.30
4B	Giving an Example	.56	.77	1.34	.36	.97	.29	.74	.78	1.41
4C1	Giving an Explanation on Topic	1.04	.43	1.05	1.64	2.65	.32	.80	1.61	2.59
4C2	Giving an Explanation off Topic	.00	0	0	0	0	0	0	0	0
4D1	Giving Directions on Topic	8.22	9.13	5.30	7.41	4.84	9.13	5.46	7.59	4.78
4D2	Giving Directions off Topic	1.81	1.62	2.09	2.01	2.18	1.76	1.68	1.86	2.45
4E	Summarizing	.81	.60	1.19	1.02	1.75	.40	.77	1.14	1.84
4F	Giving Application	.25	.36	.92	.14	.41	.13	.41	.34	.87
4G	Contrasting & Comparing	.29	.40	.86	.19	.45	.30	.83	.29	.57
4H	Direct Cueing	1.06	1.62	2.16	.52	.94	1.34	1.93	.86	1.57
5A	Asking for Answer	23.82	24.46	12.71	23.43	11.42	24.04	10.47	23.88	13.23
5BA	Asking for Additional Information	4.94	6.12	4.57	3.81	3.72	5.04	4.03	4.90	4.54
5BD	Asking for Development of Response	2.76	3.72	3.63	1.83	2.27	3.11	2.92	2.51	3.34
5BP	Asking for an Application of Response	.38	.63	1.43	.13	.49	.29	.87	.45	1.24
5C	Asking for Agreement	.77	.53	1.54	1.02	1.85	.66	1.67	.87	1.75
5D	Asking for Disagreement	.13	.14	.58	.13	.37	.06	.24	.19	.61
5E	Asking for a Judgment	.51	.68	1.40	.36	.80	.44	.90	.58	1.31
5F	Asking for Performance Appraisal, Testing	.15	.18	.51	.12	.44	.17	.49	.14	.46
5G	Asking if Information Understood	1.34	1.01	1.32	1.69	2.36	.89	1.11	1.71	2.34

Table 14

Mean Observed Frequencies of Teacher Instruction Categories:
Socioeconomic Subgroups

Category No.	Category Name	Lower Class Schools N=20 M.	S.D.	Lower Middle Class Schools N=54 M.	S.D.	Middle Class Schools N=70 M.	S.D.	Upper Middle Class Schools N=42 M.	S.D.
1	Gaining Attention	.80	1.20	.54	.57	.69	.81	.55	.63
2	Goal Statement	.25	.44	.46	.50	.46	.56	.33	.53
3	Stimulating Recall	.35	.59	.22	.46	.33	.61	.33	.48
4A1	Stating Fact on Topic	12.20	6.86	14.39	6.96	14.86	6.03	15.21	8.67
4A2	Stating Fact off Topic	.70	.93	.43	.90	.60	1.00	.79	1.57
4B	Giving an Example	.60	1.23	.46	.93	.54	1.02	.71	1.67
4C1	Giving an Explanation on Topic	.25	.72	1.17	2.35	1.16	1.89	1.05	2.48
4C2	Giving an Explanation off Topic	0	0	0	0	0	0	0	0
4D1	Giving Directions on Topic	9.65	4.07	7.48	4.66	8.37	5.20	8.45	6.00
4D2	Giving Directions off Topic	2.25	2.84	1.33	1.58	1.74	1.81	2.36	2.74
4E	Summarizing	.40	.94	.80	1.46	1.00	1.61	.71	1.61
4F	Giving Application	.20	.62	.37	.98	.26	.67	.10	.30
4G	Contrasting & Comparing	.35	.67	.33	.87	.33	.63	.17	.54
4H	Direct Cueing	1.00	2.13	1.22	2.03	1.23	1.73	.64	1.06
5A	Asking for Answer	21.35	9.58	24.28	13.48	25.19	12.31	22.69	10.79
5BA	Asking for Additional Information	5.60	5.34	5.02	4.38	5.23	4.20	4.14	3.89
5BD	Asking for Development of Response	3.00	4.30	2.78	3.46	2.97	2.95	2.33	2.50
5BP	Asking for an Application of Response	.75	2.10	.46	1.22	.33	.76	.19	.59
5C	Asking for Agreement	1.15	2.32	.98	1.93	.57	1.36	.67	1.62
5D	Asking for Disagreement	.40	.82	.09	.35	.11	.53	.10	.30
5E	Asking for a Judgment	.35	.74	.46	.97	.46	1.03	.76	1.62
5F	Asking for Performance Appraisal, Testing	0	0	.09	.29	.31	.69	.02	.15
5G	Asking if Information Understood	.85	.74	1.33	2.45	1.46	1.84	1.43	1.76

Teacher Feedback Lexicon. What feedback techniques are commonly used by teachers? Table 15 contains information bearing on this question. The most common reaction to positive achievement in the classrooms is to inform the pupil simply that he is right, without much embellishment (A1A). General praise (A1D1) is a distant second in frequency as a response to positive behavior. Evidently behavior modification techniques have not been widely adopted. Feedback in the form of granting privileges, giving tokens, etc., and using social reinforcement (A1E) is unusual.

When there are both right and wrong elements in a child's answer to a problem or a question, it is common for the teacher to attend to the negative aspect of the response (A2B3) or to simply acknowledge hearing the response (A2A). In giving feedback teachers also frequently attend to only the positive aspects of the response (A2B2), and often ask the child to elaborate (A2C).

Reacting to poor achievement, that is to a poor answer or an incorrect answer to a question, teachers most often ask a child to try again (A3C), tell him he is wrong (A3A1), or supply him with the right answer (A3A2). Seldom did the observers see behavioral modification techniques in use (A3D2, A3D4, A3D8, A3F), and rarely did they observe harshly punitive behavior (A3D3, A3H2). Most unsettling to educational psychologists and teacher educators would be the fact that there is very little guidance given when errors are made (A3B2 and A3B1).

Table 15, 16 and 17 present data facilitating the comparison of the amount and kind of feedback given by second and fifth grade

teachers, teachers giving reading lessons and mathematics lessons, and teachers with pupils of various socioeconomic status. Judging from the information in Table 15, second grade teachers are more likely than fifth grade teachers to ignore positive achievement and to direct a child to try again, should he fail. Fifth grade teachers tend simply to acknowledge hearing pupil responses. Perhaps one of the reasons for these differences is the fact that the second grade curriculum requires pupils to respond more overtly. Teachers ask children to try again and may ignore many pupil answers in the process of eliciting the correct response.

In terms of feedback for achievement, reading and mathematics lessons seem very similar. One possible exception is in the use of verbal reprimand. Fifth grade teachers are more likely to reprimand (A3D1) a pupil for an academic failure than are second grade teachers (see also A3H1 and A3H2).

From the data in Table 17, it appears that teachers of lower class pupils seem to give more simple knowledge of results (A1A) and fewer elaborated knowledge of results (A1B) for achievement and they also use more general praise than do other teachers. They ignore positive achievement (A1F) less, perhaps because it occurs less frequently, and they more often ask pupils to elaborate when there are both positive and negative aspects to what they have done. Teachers in upper middle class schools seem to be distinctive in that some of them are more likely to reprimand (A3D1) a child for failing than are other teachers, a fact which may reflect high expectations for performance.

Table 15

Mean Observed Frequencies of Teacher Feedback Categories for Achievement: Total Sample and Grade Level Subgroups

Category No.	Category Name for Achievement	Total Sample N=186 Mean	Second Grade N=82 M.	S.D.	Fifth Grade N=104 M.	S.D.
A1A	Simple knowledge of results	65.31	70.79	30.24	60.99	29.56
A1B	Elab. knowledge of results	8.06	7.02	6.31	8.88	6.99
A1C	Directions to child to elaborate	5.69	4.48	9.66	6.65	10.21
A1D1	General verbal praise	9.62	12.88	10.11	7.06	9.05
A1D2	Granting privileges	.01	.01	.11	0	0
A1D3	Positive physical contact	.79	1.10	2.42	.55	1.59
A1D4	Award of consumables	.52	.82	2.60	.29	2.08
A1E	Teacher directs others to give feedback	.24	.16	.46	.31	1.09
A1F	Ignoring of positive behavior or achievement	1.43	2.22	4.50	.81	2.36
A2A	Acknowledges hearing response	19.22	13.34	18.40	23.85	21.38
A2B1	Indication of both positive & negative	5.51	5.90	9.62	5.20	8.96
A2B2	Both but emphasis on positive	11.12	12.46	11.80	10.07	9.68
A2B3	Both but emphasis on negative	19.31	18.02	10.90	20.32	12.00
A2C	Direction to child to elaborate	9.05	9.55	9.20	8.66	11.66
A2D	Promise of reward contingent on change	.16	.24	.70	.10	.33
A2E	Teacher directs others	1.84	1.62	3.24	2.02	3.62
A2F	Ignores	1.94	2.04	2.46	1.86	3.08
A3A1	Answer termed wrong	6.55	6.67	6.58	6.45	6.70
A3A2	Correct answer given	6.76	6.94	8.16	6.62	8.86
A3B1	Error elaborated	2.18	1.77	3.00	2.50	4.54
A3B2	Correct answer elaborated	.75	.52	1.00	.92	1.45
A3C	Direction to try again	7.69	9.96	8.21	5.90	6.76
A3D1	Verbal reprimand	1.05	1.65	6.17	.58	1.10
A3D2	Withdrawal of priviliges	.06	.06	.29	.06	.23
A3D3	Aversive physical contact	.06	.08	.50	.04	.24
A3D4	Denial of consumables, tokens	.01	.01	.11	.01	.10
A3D5	Directions to stop	.18	.23	.63	.14	.59
A3D6	Warning of consequences	.23	.26	1.11	.21	.68
A3D7	Directions to proceed	.52	.89	2.21	.23	.64
A3D8	Isolation	.03	.04	.25	.03	.22
A3E	Others	.80	.76	1.84	.84	1.97
A3F	Ignores	.20	.29	.66	.12	.39
A3G	Defeat	.00	0	0	0	0
A3H1	Acceptance of DK	1.75	1.42	2.70	2.01	3.06
A3H2	Rejection of DK	.33	.35	.97	.32	1.04

Table 16

Mean Observed Frequencies of Teacher Feedback Categories for Achievement: Reading and Math Subgroups

Category No.	Category Name for Achievement	Reading N=93 M.	Reading N=93 S.D.	Math N=93 M.	Math N=93 S.D.
A1A	Simple knowledge of results	65.59	29.24	65.02	31.24
A1B	Elab. knowledge of results	8.98	6.66	7.14	6.74
A1C	Directions to child to elaborate	6.87	10.79	4.52	9.06
A1D1	General verbal praise	9.77	10.75	9.47	9.11
A1D2	Granting privileges	.10	.10	0	0
A1D3	Positive physical contact	.37	.86	1.22	2.66
A1D4	Award of consumables	.29	.76	.75	3.20
A1E	Teacher directs others to give feedback	.23	.80	.26	.94
A1F	Ignoring of positive behavior or achiev.	2.09	4.29	.77	2.41
A2A	Acknowledges hearing response	18.19	19.58	20.24	21.89
A2B1	Indication of both positive & negative	4.94	7.88	6.09	10.44
A2B2	Both but emphasis on positive	10.45	9.66	11.80	11.63
A2B3	Both but emphasis on negative	17.36	10.19	21.25	12.53
A2C	Direction to child to elaborate	9.12	10.55	8.99	11.12
A2D	Promise of reward contingent on change	.09	.38	.24	.63
A2E	Teacher directs others	1.55	3.35	2.14	3.54
A2F	Ignores	1.87	2.87	2.00	2.77
A3A1	Answer termed wrong	6.18	6.90	6.91	6.36
A3A2	Correct answer given	7.96	9.82	5.56	6.88
A3B1	Error elaborated	1.98	4.38	2.38	3.46
A3B2	Correct answer elaborated	.83	1.51	.68	1.01
A3C	Direction to try again	7.69	8.16	7.70	7.21
A3D1	Verbal reprimand	.52	1.16	1.58	5.82
A3D2	Withdrawal of privileges	.06	.25	.05	.27
A3D3	Aversive physical contact	.05	.31	.06	.44
A3D4	Denial of consumables, tokens	.01	.10	.01	.10
A3D5	Directions to stop	.16	.68	.19	.54
A3D6	Warning of consequences	.16	.61	.30	1.10
A3D7	Directions to proceed	.33	1.09	.71	1.93
A3D8	Isolation	.04	.25	.02	.21
A3E	Others	.94	2.16	.67	1.61
A3F	Ignores	.22	.59	.18	.46
A3G	Defeat	0	0	0	0
A3H1	Acceptance of DK	2.28	3.26	1.22	2.43
A3H2	Rejection of DK	.22	.76	.45	1.19

Table 17

Mean Observed Frequencies of Teacher Feedback Categories for Achievement: Socioeconomic Subgroups

Category No.	Category Name for Achievement	Lower Class N=20		Lower Middle N=54		Middle Class N=70		Upper Middle N=42	
		M.	S.D.	M.	S.D.	M.	S.D.	M.	S.D.
A1A	Simple knowledge of results	77.65	24.03	69.06	27.47	61.27	28.81	61.33	36.50
A1B	Elab. knowledge of results	6.90	4.98	7.48	6.06	9.16	7.61	7.52	6.76
A1C	Directions to child to elaborate	6.95	12.10	4.24	8.14	7.59	11.81	3.81	7.00
A1D1	General verbal praise	14.85	15.28	8.57	8.83	9.00	8.94	9.52	9.32
A1D2	Granting privileges	0	0	0	0	.01	.12	0	0
A1D3	Positive physical contact	.95	1.50	.09	.45	1.14	2.50	1.02	2.35
A1D4	Award of consumables	.50	.95	.65	9.92	.23	.80	.86	3.43
A1E	Teacher directs others to give feedback	.15	.49	.20	.76	.36	1.19	.14	.35
A1F	Ignoring of positive behavior or achiev.	.50	.89	.70	1.70	1.66	4.15	2.43	4.58
A2A	Acknowledges hearing response	17.60	21.29	15.17	15.02	21.19	23.31	21.69	22.02
A2B1	Indication of both positive & negative	4.75	3.54	6.72	10.29	6.39	11.27	2.86	4.23
A2B2	Both but emphasis on positive	13.60	10.65	11.09	9.25	12.44	13.21	7.79	6.36
A2B3	Both but emphasis on negative	18.80	10.12	22.63	13.64	18.46	10.52	16.69	10.23
A2C	Direction to child to elaborate	13.00	13.42	12.15	12.76	7.20	9.30	6.29	7.33
A2D	Promise of reward contingent on change	.15	.67	.17	.42	.11	.36	.24	.76
A2E	Teacher directs others	2.10	3.71	1.61	3.28	1.09	3.16	2.29	4.03
A2F	Ignores	2.60	2.82	2.24	3.12	1.79	2.84	1.48	2.3
A3A1	Answer termed wrong	7.15	9.76	6.98	6.13	6.36	5.55	5.69	7.23
A3A2	Correct answer given	7.00	7.12	7.87	9.18	6.57	8.89	5.52	7.50
A3B1	Error elaborated	1.45	1.67	2.57	5.46	2.41	3.47	1.62	3.02
A3B2	Correct answer elaborated	.70	1.26	.59	.84	.77	1.18	.93	1.84
A3C	Direction to try again	7.00	5.50	9.26	8.55	6.93	6.94	7.29	6.50
A3D1	Verbal reprimand	.50	.76	1.37	3.77	.33	.91	2.10	7.61
A3D2	Withdrawal of privileges	.05	.22	.07	.26	.04	.17	.10	.37
A3D3	Aversive physical contact	0	0	.04	.27	.01	.12	.19	.81
A3D4	Denial of consumables, tokens	0	0	.02	.14	0	0	.02	.15
A3D5	Directions to stop	.20	.52	.17	.47	.10	.42	.31	.95
A3D6	Warning of consequences	.15	.37	.44	1.46	.10	.42	.21	.66
A3D7	Directions to proceed	.70	2.16	.46	1.21	.33	.76	.83	2.44
A3D8	Isolation	0	0	0	0	.03	.24	.10	.37
A3E	Others	.95	2.67	.65	1.40	.87	1.96	.81	2.02
A3F	Ignores	.65	.99	.11	.32	.19	.49	.12	.40
A3G	Defeat	0	0	0	0	0	0	0	0
A3H1	Acceptance of DK	1.90	2.17	1.22	2.04	1.96	3.64	2.00	2.80
A3H2	Rejection of DK	.45	.95	.50	1.18	.20	.67	.29	1.24

In looking at Tables 18, 19 and 20 we find frequencies of feedback given to pupil classroom behavior which is nonachievement related. These frequencies reflect the teacher's style of enforcing classroom discipline and decorum. From Table 18 it is clear that although teachers do tell pupils that they are correct when they are behaving well (B1A), they are much more likely to react or attend to negative behavior (B2B3, B3D1). Some teachers do attend to the positive when both positive and negative behaviors occur simultaneously (B2B2). When clearly negative behavior occurs the usual reaction is a verbal reprimand (B3D1), directions to proceed with an assigned task (B3D7), or directions to stop the behavior (B3D5). Very few teachers purposefully ignore negative behavior (B3F), and very seldom does a teacher involve peers in giving feedback to children whose behavior was inappropriate (B3E, B2E). Behavioral modification techniques in the form of using privileges (B1D2, B3D2), and the use of tokens (B1D4, B3D4) are also infrequent in these classrooms.

Again there are few differences between second and fifth grade classrooms, the only notable one being the fact that misbehaving second grade pupils are told more often to proceed with an assigned task (B3D7) than are fifth grade children. Perhaps this may be attributed to the notion that second graders are more distractable than fifth graders and may need more direction. The fifth grader's misbehavior may be more willful.

Table 19 contrasts frequency of feedback during reading lessons and mathematics lessons. Here there are no discrepancies worthy of mention. Misbehavior is reacted to in the same way, whether in a reading lesson or in a mathematics lesson.

Table 20 contrasts frequency of feedback events across socio-economic groups. There are more feedback events of all types observed in the tapes of teachers in lower class schools, reflecting the possibility that there is more nonachievement oriented behavior in these classrooms. There is more praise and recognition for good behavior in the lower SES classrooms (B1C, B1A), but there is also more attention to negative aspects of behavior (B2B3) and more direction to stop negative behavior (B3D5) and to get on with an assigned task (B3D7).

The information contained in the preceding paragraphs is difficult to evaluate. While it gives us information about what is observable on videotapes, it leaves us with a dilemma. The fact that a particular event is infrequently observed does not mean that the event is not an important behavior. One teacher may only give one child positive physical contact during a lesson, whereas another teacher may give several children a pat or an embrace. Nevertheless, one significant contact between teacher and child in one class may have been far more important than numerous physical contacts between another teacher and his/her pupils. Moreover, some forms of feedback or events of instruction are only appropriate at very few times. In many cases, the vary fact itself that a teacher exhibits a particular behavior is more important than the number of times he/she performs that behavior. The frequency data, then, must be used with great care in designing future experimental research.

Table 18

Mean Observed Frequencies of Teacher Feedback Categories for Behavior: Total Sample and Grade Level Subgroups

Category No.	Category Name for Behavior	Total Sample N=186	Second Grade N=82		Fifth Grade N=104	
		Mean	M.	S.D.	M.	S.D.
B1A	Simple knowledge of results	2.04	2.71	5.03	1.51	2.37
B1B	Elab. knowledge of results	.08	.10	.34	.07	.25
B1C	Directions to child to elaborate	.03	.07	.31	0	0
B1D1	General verbal praise	.91	1.28	2.72	.64	1.37
B1D2	Granting privileges	.01	0	0	.01	.10
B1D3	Positive physical contact	.27	.45	1.35	.12	.36
B1D4	Award of consumables	.21	.45	1.78	.03	.17
B1E	Teacher directs others to give feedback	.02	.04	.33	0	0
B1F	Ignoring of positive behavior or achiev.	.02	.02	.16	.01	.10
B2A	Acknowledges hearing response	2.04	2.50	7.12	1.67	3.24
B2B1	Indication of both positive & negative	1.30	1.99	4.73	.76	1.57
B2B2	Both but emphasis on positive	1.83	2.38	4.43	1.40	2.55
B2B3	Both but emphasis on negative	7.41	8.45	8.41	6.59	6.54
B2C	Direction to child to elaborate	.63	.83	1.98	.47	1.37
B2D	Promise of reward contingent on change	.04	.06	.24	.02	.14
B2E	Teacher directs others	.03	.05	.31	.01	.10
B2F	Ignores	.20	.37	1.03	.07	.35
B3A1	Answer termed wrong	.20	.28	.86	.14	.40
B3A2	Correct answer given	.08	.06	.24	.09	.34
B3B1	Error elaborated	.02	.02	.16	.01	.10
B3B2	Correct answer elaborated	.02	0	0	.04	.19
B3C	Direction to try again	.07	.11	.42	.03	.22
B3D1	Verbal reprimand	5.18	4.59	5.24	5.65	6.34
B3D2	Withdrawal of privileges	.11	.13	.60	.10	.30
B3D3	Aversive physical contact	.17	.22	.63	.12	.50
B3D4	Denial of consumables, tokens	.00	0	0	0	0
B3D5	Directions to stop	1.96	2.29	3.32	1.70	3.15
B3D6	Warning of consequences	.50	.55	1.07	.46	.87
B3D7	Directions to proceed	2.22	2.93	4.55	1.67	2.89
B3D8	Isolation	.50	.39	.86	.58	1.68
B3E	Others	.02	.04	.19	.01	.10
B3F	Ignores	.16	.24	.78	.10	.36
B3G	Defeat	.00	0	0	0	0
B3H1	Acceptance of DK	.02	.04	.25	0	0
B3H2	Rejection of DK	.01	.01	.11	.01	.10

Table 19

Mean Observed Frequencies of Teacher Feedback Categories for Behavior: Reading and Math Subgroups

Category No.	Category Name for Behavior	Reading N=93		Math N=93	
		M.	S.D.	M.	S.D.
B1A	Simple knowledge of results	2.38	4.52	1.70	2.93
B1B	Elab. knowledge of results	.06	.29	.10	.30
B1C	Directions to child to elaborate	.04	.25	.02	.15
B1D1	General verbal praise	1.04	2.54	.89	1.52
B1D2	Granting privileges	0	0	.01	.10
B1D3	Positive physical contact	.38	1.27	.16	.42
B1D4	Award of consumables	.30	1.59	.12	.61
B1E	Teacher directs others to give feedback	.03	.31	0	0
B1F	Ignoring of positive behavior or achiev.	.02	.15	.01	.10
B2A	Acknowledges hearing response	2.58	6.62	1.50	3.52
B2B1	Indication of both positive & negative	1.36	2.79	1.25	3.92
B2B2	Both but emphasis on positive	2.01	3.80	1.66	3.24
B2B3	Both but emphasis on negative	8.08	7.82	6.74	7.06
B2C	Direction to child to elaborate	.60	1.27	.66	2.00
B2D	Promise of reward contingent on change	.04	.20	.03	.18
B2E	Teacher directs others	.02	.21	.03	.23
B2F	Ignores	.28	.96	.12	.41
B3A1	Answer termed wrong	.14	.38	.26	.83
B3A2	Correct answer given	.05	.31	.10	.30
B3B1	Error elaborated	.03	.18	0	0
B3B2	Correct answer elaborated	.02	.15	.02	.15
B3C	Direction to try again	.04	.29	.09	.35
B3D1	Verbal reprimand	4.87	5.72	5.50	6.07
B3D2	Withdrawal of privileges	.14	.58	.09	.28
B3D3	Aversive physical contact	.25	.69	.09	.38
B3D4	Denial of consumables, tokens	0	0	0	0
B3D5	Directions to stop	2.08	3.58	1.85	2.86
B3D6	Warning of consequences	.46	.95	.54	.97
B3D7	Directions to proceed	2.10	3.73	2.36	3.90
B3D8	Isolation	.39	.83	.60	1.76
B3E	Others	.03	.18	.01	.10
B3F	Ignores	.19	.71	.13	.42
B3G	Defeat	0	0	0	0
B3H1	Acceptance of DK	.03	.23	0	0
B3H2	Rejection of DK	.01	.10	.10	.10

Table 20

Mean Observed Frequencies of Teacher Feedback Categories for Behavior: Socioeconomic Subgroups

Category No.	Category Name for Behavior	Lower Class N=20		Lower Middle N=54		Middle Class N=70		Upper Middle N=42	
		M.	S.D.	M.	S.D.	M.	S.D.	M.	S.D.
B1A	Simple knowledge of results	2.65	3.59	1.76	2.40	2.51	4.99	1.31	3.03
B1B	Elab. knowledge of results	.15	.37	.07	.33	.09	.28	.05	.22
B1C	Directions to child to elaborate	.10	.45	0	0	.03	.17	.05	.22
B1D1	General verbal praise	2.35	4.48	.48	1.00	.99	2.00	.69	1.07
B1D2	Granting privileges	0	0	0	0	0	0	.02	.15
B1D3	Positive physical contact	.55	.76	.11	.46	.33	1.37	.24	.53
B1D4	Award of consumables	.80	2.07	.02	.14	.27	1.57	.10	.37
B1E	Teacher directs others to give feedback	.15	.67	0	0	0	0	0	0
B1F	Ignoring of positive behavior or achiev.	.05	.22	0	0	0	0	.05	.22
B2A	Acknowledges hearing response	2.50	6.84	.96	2.26	2.89	7.09	1.79	3.42
B2B1	Indication of both positive & negative	1.20	1.80	1.33	2.50	1.84	4.90	.40	.88
B2B2	Both but emphasis on positive	2.85	5.13	1.50	2.74	2.33	4.20	.95	1.64
B2B3	Both but emphasis on negative	11.15	8.06	7.35	8.18	7.46	7.26	5.62	5.95
B2C	Direction to child to elaborate	.65	.99	.89	1.97	.77	1.99	.05	.22
B2D	Promise of reward contingent on change	.10	.31	.04	.19	.03	.17	.02	.15
B2E	Teacher directs others	.10	.45	.02	.14	.03	.24	0	0
B2F	Ignores	.65	1.23	.26	1.01	.07	.31	.12	.40
B3A1	Answer termed wrong	.65	1.46	.20	.53	.09	.28	.17	.54
B3A2	Correct answer given	.20	.52	.13	.39	.04	.20	0	0
B3B1	Error elaborated	.05	.22	0	0	.01	.12	.02	.15
B3B2	Correct answer elaborated	0	0	0	0	.01	.12	.07	.26
B3C	Direction to try again	.15	.22	.15	.53	.03	.17	.02	.15
B3D1	Verbal reprimand	5.65	4.65	4.89	5.33	5.11	6.08	5.45	6.88
B3D2	Withdrawal of privileges	.10	.31	.06	.23	.17	.66	.10	.30
B3D3	Aversive physical contact	.40	.94	.11	.50	.16	.53	.14	.42
B3D4	Denial of consumables, tokens	0	0	0	0	0	0	0	0
B3D5	Directions to stop	2.50	3.02	2.13	3.02	1.84	3.75	1.69	2.67
B3D6	Warning of consequences	.60	1.23	.48	.93	.63	1.04	.26	.67
B3D7	Directions to proceed	4.25	6.57	1.98	3.07	1.79	2.88	2.31	4.11
B3D8	Isolation	.90	2.47	.26	.62	.43	1.35	.71	1.42
B3E	Others	.05	.22	0	0	.04	.20	0	0
B3F	Ignores	.30	.73	.07	.33	.23	.77	.07	.34
B3G	Defeat	0	0	0	0	0	0	0	0
B3H1	Acceptance of DK	0	0	.06	.30	0	0	0	0
B3H2	Rejection of DK	0	0	.02	.14	.01	.12	0	0

Data Reduction

Because there are so many different variables which were observed with a low frequency of occurrence, and because one of the sources of unreliability is observers categorizing a given event in two similar but different categories, it seemed reasonable to collapse categories and add similar low frequency events together. Both the Teacher Instruction events and the Teacher Feedback events were logically reevaluated. Event frequencies were combined into larger categories and the reliabilities were recomputed using Spearman's r to estimate the correlation between two observers.

Teacher Instruction events. The Teacher Instruction categories have been reduced from 22 separate events to six categories: Introductory Remarks, Expository Remarks, Giving Directions, Asking General Questions, Asking Prodding Questions or Cueing, and Asking Questions to Determine if Pupil has Understood Lesson. The six categories, the events combined to make up these categories, and the calculated reliabilities are to be found in Table 21. The Spearman rank order correlations range from .53 to .85. Not all of the original 22 categories remained in the system. Events 4A2 (Stating Fact off Topic), 4D2 (Giving Directions off Topic), 5D (Asking for Disagreement), and 5E (Asking for a Judgement) are infrequently observed and seem less logically connected to the other events. They were dropped from consideration.

Teacher Feedback events. Table 22 presents the Teacher Feedback categories after the events were combined, along with the inter-rater reliability. Observers had recorded with the event whether or not the feedback event was directed at pupil achievement

or at pupil behavior. Those events directed at achievement were separated from those oriented to pupil behavior. At first, we examined just the events which were connected to pupil achievement oriented behavior. Four events were maintained in their original form, category 1A, 1B, 1D3, and 3C. They were fairly reliable to begin with, and when the events associated with classroom management of behavior were deleted, the reliability increased.

One reason observers may have confused events was the subtle distinction between type 1 events (Reactions to Positive Achievement) and type 2 events (Reactions to Mixed Achievement). Type 1 and type 2 events were combined for 1C and 2C, 1E and 2E, 1F and 2F. Because of the small frequencies, all of the remaining type 2 events were combined, and all of the error feedback events were placed in one category.

Next we turned our attention to the Teacher Feedback events which are teacher reactions to nonachievement oriented pupil behavior. Only those events which involved praise or punishment were present in sufficient frequency to justify combining them. Consequently the four different forms of praise were collapsed together (1D events), the seven different events regarding punishment for disruptive behavior were combined, and the two events involved with warning pupils about negative behavior were grouped together.

The Spearman correlation coefficients in Table 22 range from .44 to .82. Although there are two coefficients below .50, we decided to keep them rather than reduce the data further. The 35 original events (70 if achievement and behavior events were considered separately) were reduced to 12 categories.

Rating scales. To reduce the amount of data on the rating scales, the first step was to submit the reliable items from the ratings to factor analysis. Accordingly, the investigator read the data from the eight reliable items on the Observation Classroom Summary Record into the Kaiser Little Jiffy, Mark IV, Exploratory Factor Analysis program and obtained the factor structure matrix in Table 23. The three factors which emerged seemed easily labeled: Praise, Need for Intervention, and Students Initial Interaction. The eight items could, then, be reduced further into three factor scores.

The same procedure was attempted with the reliable items from the two Videotape Rating Schedules. The resulting factor matrices were not easily interpreted. As a result we decided to use the 11 reliable items individually and not combine them in any way.

Table 21

Combined Categories for the
Teacher Instruction Lexion

New Category Label		Events Combined	Spearman's r Between Two Observers N = 30 Tapes
Introductory remarks	1	Gaining attention	
	2	Goal statement	.61
	3	Stimulating recall	
Expository remarks	4A1	Stating facts	.85
	4B	Giving an example	
	4E	Summarizing	
	4F	Giving application	
	4G	Contrasting and comparing	
Giving directions	4D1	Giving directions	.53
	4C	Giving an explanation	
Asking general questions	5A	Asking for answer	.83
	5F	Asking for performance appraisal, Testing	
Asking prodding questions or cueing	5BA, D & P	Asking followup questions	.80
	4H	Direct cueing	
Asking questions to determine if pupil has understood lessons	5G	Asking if information understood	.76
	5C	Asking for agreement	

Table 22

Combined Categories for the
Teacher Feedback Lexion

New Category Label	Events Combined	Spearman's r between two observers N = 31 Tapes
Simple knowledge of results	1A	.82
Elaborated knowledge of results	1B	.51
Child asked to elaborate answer	1C + 2C	.61
Positive Physical contact	1D3	.71
Encouraging other pupils to give feedback	1E + 2E	.72
Ignoring positive or mixed achievement	1F + 2F	.46
Weak acknowledgement of achievement	2A + 2B1 + 2B2 + 2B3	.44
Calling attention to error	3A1 + 3A2 + 3B1 + 3B2	.63
Pupil asked to try again	3C	.51
Praise for good behavior	1D1 + 1D2 + 1D3 + 1D4	.53
Punishment for disruptive behavior	3D1 + 3D2 + 3D3 + 3D4 + 3D5 + 3D7 + 3D8	.74
Warning for negative behavior	3D6 + 2D	.72

Table 23

Factor Structure Matrix of Reliable Observation Summary Record Items

		Praise, Factor 1	Need for Intervention, Factor 2	Students Initiate Interaction, Factor 3
1.	Frequent Praise	.87	-.31	-.40
2.	Praise to many	.88	-.36	-.39
3.	Students initiate interaction	-.28	.12	.70
4.	Adults initiate interaction	.36	-.27	-.72
13.	Students in control	.40	-.78	-.29
14.	Need for intervention	-.27	.83	.19
20.	Students require motivation	-.27	.79	.20
24.	Adults negative	-.64	.59	.31

Table 24

Correlation Coefficients Between Teacher Instruction Behavior Events and Spring Test Scores Partialling Out Fall Test Scores

	California Achievement Test Reading Comprehension		Reading Attitude Measure		California Achievement Test Mathematics Concepts		California Achievement Test Mathematics Computation		Mathematics Attitude Measure	
	2nd	5th	2nd	5th	2nd	5th	2nd	5th	2nd	5th
Introductory Remarks	-.11	-.08	-.06	.11	.19	.00	-.19	.00	.07	.04
Expository Remarks	.11	.05	-.17	.23*	-.23*	-.06	.15	-.07	.04	-.09
Giving Directions	-.06	-.20*	-.02	.03	.06	-.33**	.08	-.36**	.05	-.39**
Asking General Questions	.22*	.03	.14	.06	-.13	-.19*	-.02	-.22*	.11	-.20*
Asking Prodding Questions or Cueing	.02	-.04	.02	-.05	.09	-.03	-.01	-.02	.08	.01
Asking Questions to Determine if Pupil Has Understood Lesson	.09	.05	.00	-.07	.17	-.03	.09	.02	.14	.09

* $p \leq .10$

** $p \leq .05$

Relationship Between Information Collected from Videotapes and Pupil Achievement

One of the goals of this phase of the BTES project has been to isolate for further study teacher behaviors which may be related to pupil growth. Two techniques have been used to determine the relationship between the teacher behaviors obtained from the videotapes and pupil achievement. The first method consisted of correlating the videotape-derived observational variables with the mean adjusted spring achievement and attitude measures across the total group of second and fifth grade classrooms. The second method involved contrasting the five most effective teachers (effective defined as most pupil growth) and the five least effective teachers at each grade level and in each subject area.

One caution should be noted at this point. We are correlating variables derived from the viewing of one 40-minute segment of videotape collected at one time during the school year with measures of pupil growth across the academic year. To expect any meaningful relationships to emerge from such a limited sample of a teacher's behavior is extremely optimistic. Any relationships which do turn up should be considered very seriously as important aspects of teacher behavior to be furtner examined or explained.

Partial correlation is a statistical technique which allows the researcher to calculate a measure of association between two variables while adjusting for the effects of one (or more) additional or control variables. In this instance, we are interested in the relationship of videotape observational variables to pupil performance in the spring, correcting for or adjusting to the same pupil's performance in the fall. Table 24 presents the partial correlation

coefficients of the Teacher Instruction measures with the appropriate criterion variables from those collected by the BTES project. For this analysis, observational variables collected during reading lessons were related to the California Achievement Test (CAT)-Reading Comprehension* and to the ETS developed Reading Attitude measure. The observational variables obtained from the mathematics lesson tapes were related to the CAT-Mathematics Concepts, the CAT-Mathematics Computation, and the ETS developed Mathematics Attitude measure.

An examination of Table 24 indicates that three of the Teacher Instruction Behavior categories are not related to achievement or attitude toward subject matter at either the second or fifth grade level: Giving Introductory Remarks, Asking Prodding Questions or Cueing, and Asking Questions to Determine if the Pupil has Understood the Lesson. However, the other three categories in Table 24 are related.

During reading lessons at the fifth grade level, the more expository remarks made by the teacher, the better is the pupil's attitude toward reading. Evidently, the more the teacher talks about the reading assignment and gives facts, examples, summaries, and applications, etc., the better the pupils feel about reading and reading lessons. On the other hand, at the second grade, the more a teacher makes expository remarks, the less well the pupils do at

---

*Other achievement tests were administered to the project children. In the interest of reducing the number of variables we selected the CAT tests because they are the most completely studied. Also, these test results were the first available and project deadlines precluded the use of the ETS/Calfee tests.

acquiring mathematical concepts. Perhaps less talking and more practice or manipulation is related to the acquisition of mathematical concepts; listening to the teacher and working with materials may be mutually exclusive in a second grade classroom.

Giving Directions is negatively correlated with both achievement and attitude in almost all cases at the fifth grade level. The more directions a teacher gives, the less well the class, as a group, performs. A correlation coefficient does not imply causality; we might also say the less well the class does, the more directions a teacher is likely to give. Giving a great many directions may be a sign that the class is not working at the correct level. If a teacher must continually give directions and explanations, the pupils probably are not able to do the work on their own. Perhaps better teachers do not need to give directions or explanations more than once.

Asking General Questions is negatively related also to mathematics achievement and attitude toward mathematics at the fifth grade level. The more verbal behavior in the form of questioning (and in mathematics this would probably be in the form of asking children to recite answers) at the fifth grade level, the poorer the pupils' achievement and the more negative the pupils' attitude toward mathematics. If drill and poor achievement are related, one would expect the more covert the practice and the more individualized the instruction, the better the achievement. This result should be explored further.

Asking Questions is positively related to gains in reading comprehension at the second grade level, however. This seems

reasonable, for the technique of questioning children about what they have read is a standard one for stimulating the discussion which fosters comprehension. More research should be done to determine the nature and timing of questions which are most conducive to the development of reading comprehension.

Teacher Feedback Lexicon. Table 25 contains the correlations between the Teacher Feedback event categories and the pupil growth measures. Only one of the variables, Calling Attention to Error, is unrelated to any of the criterion measures.

Simple Knowledge of Results during reading lessons is negatively related to reading comprehension at the fifth grade. The more pupils are told they are correct when they are, the less well they do on the measure of reading comprehension. This result is a puzzling one. Perhaps pupils improve in comprehension more in the fifth grade if they are given questions which do not involve right and wrong answers. It may be that less effective teachers tend to avoid asking questions which tap pupil opinion and interpretation of reading materials and prefer simple factual recall questions to which they give simple knowledge of results. This finding might be explored further.

Elaborated Knowledge of Results correlates negatively with the measure of attitude toward mathematics. Teachers who give detailed information to pupils about correct answers do not seem to positively influence their fifth graders' attitudes.

Asking a Child to Elaborate after he has responded seems to be related to reading comprehension at the second grade level. This technique of bringing a child out and having him talk more about his

thoughts is evidently an effective one for raising achievement in reading comprehension.

Positive Physical Contact is also effective with second grade children. In the present study it was significantly corelated with reading attitude and mathematics achievement. Younger children can be encouraged to perform by teachers who are warm and "physical", but it appears that physical contact does not have much of an influence on older children.

Another teaching technique positively related to achievement in the second grade is using other pupils to give praise, encouragement and knowledge of results to children. In Table 25, Encouraging Other Pupils to Give Feedback is positively related to both reading and mathematics achievement, as well as attitude toward mathematics.

At the fifth grade, however, having other pupils give feedback is negatively related to reading achievement. Having others respond to your answers or products during a reading lesson may be embarrassing and create a climate unconducive to learning.

Ignoring Postive or Mixed Achievement may be thought of in two ways. A teacher who ignores a pupil's achievement may be communicating a lack of interest in the product, creating an unpleasant feeling in a pupil, or it may be communicating a set of high standards and expectations which would stimulate a pupil onward. At the fifth grade, this variable is negatively related to attitude toward reading. Perhaps fifth graders interpret ignoring as disinterest. In the second grade ignoring was positively correlated with mathematics achievement. Perhaps at the second grade, ignoring may be related to high expectations for the pupil. These explanations are worthy of testing in future investigations.

The variable Weak Acknowledgement of Achievement is a measure of how much a teacher reacts to a pupil achievement that is neither completely right nor wrong. Reacting to a mixed pupil response with some acknowledgement is positively correlated with achievement in both reading and mathematics for second graders. The younger pupils seem to profit from having a reactive, responsive teacher. This variable is negatively correlated with mathematics attitude in fifth grade pupils, however. Perhaps the fifth grade pupils expect more encouragement and are disappointed that it is weak. Another explanation of this finding is that what is being measured is not the teacher's feedback, but the presence of mixed quality responses. If a student is performing poorly, he probably does not have a positive attitude toward mathematics.

Encouraging a Pupil To Try Again is an important teacher behavior; if the pupil doesn't attempt a task again until he gets it correct, he will not learn. Table 25 presents an interesting finding with this variable and mathematics learning in second graders. Trying Again is negatively related to achievement. Because they have been incorrect, teachers require pupils to try the task again. But, the very fact that the pupil was wrong the first time is a sign his achievement will be low. However, being encouraged to try again is positively correlated with attitude, indicating that any encouragement a teacher does give will make a pupil more at home with mathematics. This paradoxical finding should be validated in future research.

How do teachers' reactions to non-academically oriented behavior relate to achievement and pupil attitude? Praise for

Good Behavior and/or Giving Warning for Negative Behavior has little or no effect on fifth grade pupils. Praise seems to have a positive effect on second graders' reading attitude and mathematics computation skills. The more punishment the teacher gives, however, the poorer the pupils are apt to do in mathematics. It seems reasonable that the more disruptive the mathematics class, the less a pupil will learn of mathematics concepts. However, the more warning, the more a teacher has only to remind children of consequences rather than punish them, the more mathematics learning takes place. The relation of classroom management and pupil achievement must be explored fully in any study of teacher competencies.

In looking at the table as a whole, it is interesting to note that none of the event categories is significantly related to achievement or attitudes across both grade levels. Few of the event categories are related in the same way to both achievement and attitudes in a subject, and few are related in the same way to mathematics and reading. Teacher feedback behavior does not have universal effects. Rather, certain types of feedback have effects in certain situations and not in others. Any further research on feedback must take into account more carefully the context and appropriateness of the feedback behavior.

In the fifth grade, for example, none of the correlations between feedback behavior and the criterion variables are in the positive direction. The more feedback pupils are receiving in the fifth grade, the worse is their achievement and attitude. While feedback may assist second grade pupils, the presence of feedback in fifth grade classrooms seems either to alienate pupils or to be an

Table 25

Correlation Coefficients Between Teacher Feedback Behavior Event Categories and Spring Test Scores Partialling Out Fall Test Scores

	California Achievement Test Reading Comprehension		Reading Attitude Measure		California Achievement Test Mathematics Concepts		California Achievement Test Mathematics Computation		Mathematics Attitude Measure	
	2nd	5th	2nd	5th	2nd	5th	2nd	5th	2nd	5th
Simple Knowledge of Results	.08	-.27**	-.04	.00	.18	.00	-.19	-.05	.18	-.01
Elaborated Knowledge of Results	-.05	.04	-.05	.16	.09	.03	.09	-.10	.03	-.22*
Child Asked to Elaborate	.28**	-.14	.00	.04	-.01	-.08	.20	-.04	.01	.01
Positive Physical Contact	.11	.05	.22*	.14	.24*	.00	-.18	.07	.18	-.01
Encouraging other Pupils to Give Feedback	.27**	-.22*	.15	-.14	.31**	-.01	.06	.12	.30*	.05
Ignoring Positive or Mixed Achievement	.14	.10	.12	.27*	-.07	.18	.25*	.09	-.03	.05
Weak Acknowledgement of Achievement	.25*	-.16	.07	.10	-.05	-.12	.32**	-.16	-.07	-.21*
Calling Attention to Error	-.12	.00	-.12	-.11	-.10	.07	.02	.15	-.05	.16
Pupil Asked to Try Again	-.04	.01	-.16	-.13	-.04	.04	-.41**	.06	.23*	.11
Praise for Good Behavior	-.06	.08	.21*	.15	-.18	.16	.25*	.12	.15	.07
Punishment for Disruptive Behavior	.00	.03	.04	.01	-.27*	-.02	.00	.00	-.20	.00
Warning for Negative Behavior	-.16	-.18	-.08	.01	.29*	.00	.08	.08	.03	.12

*p ≤ .10
**p ≤ .05

indication that the pupils are not performing well. Pupils who are functioning well do not need feedback at the fifth grade. When children are being rewarded intrinsically by the materials and activities they are involved with, they are learning. When pupils are not achieving well, teachers may resort to more verbal, overt feedback attempting to influence learning.

Table 26 contains partial correlations between rating scale items and the five criterion variables at both the second and fifth grade levels. All of the items are related to pupil achievement or attitude except Teacher Gestures. Teacher Movement is correlated with pupil attitude toward reading in second grade classrooms. To receive a high rating on this aspect of teacher variability, teachers usually move about the room, interacting with instructional materials. Teachers who help students act out stories or who move around the classroom during reading lessons are also rated as being high in Teacher Movement. All of these behaviors tend to enhance pupil appreciation for reading.

Teacher Enthusiasm, ordinarily a powerful variable in research on teaching, is related only to attitude toward mathematics at the second grade. In this case, the more enthusiastic the teacher, the less favorable is the attitude toward mathematics. One possible explanation for this finding is that overly enthusiastic second grade teachers somehow communicate to children that a lot of energy and effort is necessary to be successful in mathematics, a message which is somehow discouraging to children.

The rating of Time Spent on Learning Activities is positively correlated with second grade attitude toward reading and mathematics

achievement. The more time spent in learning in the second grade the better the children feel about reading and the more mathematics concepts they learn. Clearly attitude and achievement are related. When teachers give sufficient time for pupils to master matierals and to succeed with them, the pupils learn and enjoy learning. A similar explanation may serve for the obtained relationship between fifth grade reading achievement and attitude and the rating of Lesson Learning Centered. Apparently fifth graders need task-orientation both to learn and to feel good about reading. Perhaps mathematics lessons and reading lessons at the second grade are all equally learning centered resulting in insignificant correlations, but when fifth grade reading lessons are focused on reading, progress occurs.

The rating of the extent to which learning activities are differentiated for pupils, i.e., the amount of observed individualized instruction, is positively related to second grade mathematics attitude. The more teachers individualize instruction in mathematics at the second grade, the more positively the pupils feel about the subject. Fifth grade pupils, however, seem to have a more negative attitude toward reading if there are differentiated reading assignments in their classrooms. At the fifth grade, pupils are much more aware of peer attitudes and those who find themselves assigned to material below grade level may feel insulted and discouraged about reading.

The Attentiveness of Children, as might be expected, is positively related to mathematics achievement in the second grade and with attitude toward mathematics in the fifth grade. In mathematics

particularly, it is important for children to attend to the lesson. Second grade children are only able to attend to lessons for a short period of time and the more successful a teacher is at engaging the children in learning, the more the pupils progress. At the fifth grade, the relation between attention and mathematics attitudes may work in either direciton. Attention may reflect positive attitudes or stimulate them.

The technique of Delay in Naming Child to Answer a Question after it has been asked seems to be an effective one in increasing reading comprehension at the fifth grade, although it does not create a positive attitude toward reading. If done sensitively, pausing before naming a child to answer stimulates all the children in the class to formulate a reply, for any one of them may be called upon to answer. This practice can, however, create tension in the classroom which may contribute to negative attitudes. At the second grade children are much more eager to be chosen and to volunteer, and are less fearful of failure so that this practice does not have a negative influence on these children.

Once a child has been called upon to answer, a teacher may wait for a reply or, if a reply is not forthcoming, move on to another pupil. The rating of the amount of Waiting for Child to Answer is negatively related to reading attitude at the second grade. A long wait may be embarrassing to pupils at the second grade or a sign that pupils do not know answers at the fifth grade.

Two important characteristics of successful counselors are warmth and honesty. Observers made judgements of these important global human characteristics displayed by the videotaped teachers.

Although these are very general and somewhat diffuse traits, perhaps some of the methodology of research on counseling can be applied to research on teachers. The results of this exploratory study indicate positive correlations between warmth and achievement in learning mathematics concepts at both grades, between honesty and achievement in mathematics computation at both grades, and between honesty and attitude toward mathematics at the fifth grade. Warm teachers do not threaten children and are accepting of children, and this disposition seems to have an effect on learning. Honest teachers give direct feedback, letting pupils know where they stand. Honest teachers do not give gratuitous praise, an attribute which older pupils may appreciate more than younger ones.

Factor I from the Classroom Observation Summary involves the frequency with which adults give praise and the number of pupils at whom the praise is directed. The factor score for these ratings is negatively related to achievement in fifth grade mathematics classrooms. Perhaps the existence of praise in these classrooms indicates a great deal of verbal discussion but very little pupil involvement in materials on their own. Giving frequent praise may also be superficial or meaningless. Pupils may react better to judiciously given earned praise, rather than unearned flattery.

Factor II, Need for Intervention, is a summary of ratings about the amount of time necessary for a teacher to control the learning environment. This factor is significantly related to mathematics achievement in both the second and fifth grades. The less control the teacher has to exercise in the second grade, the better the achievement in concept learning. The more control the teacher

Table 26

Correlation Coefficients Between Rating Scale Items and Factor Scores, and Spring Test Scores Partialling Out Fall Test Scores

	California Achievement Test Reading Comprehension		Reading Attitude Measure		California Achievement Test Mathematics Concepts		California Achievement Test Mathematics Computation		Mathematics Attitude Measure	
	2nd	5th	2nd	5th	2nd	5th	2nd	5th	2nd	5th
Teacher Movement	.10	-.16	.20*	-.11	.08	.18	.02	.16	-.07	.09
Teacher Gestures	.08	-.04	.04	-.02	.03	.01	.18	.03	.11	.06
Teacher Enthusiasm	.06	-.04	-.09	.05	.18	-.04	.13	-.06	-.23*	.05
Time Spent on Learning Activities	.14	.07	.20*	.00	.23*	.05	.12	.06	.01	.07
Lesson Learning Centered	.05	.20*	.06	.19*	-.08	.11	.06	.08	.06	.10
Differentiation of Learning Activities	.04	-.08	.15	-.19*	.07	-.01	.07	.07	.30**	.13
Attentiveness of Children	.05	-.14	-.11	-.09	.28**	.11	.31**	.03	-.03	.28**
Delay in Naming Child to Answer a Question	-.02	.25**	.23*	-.40**	-.15	.10	.04	.12	-.08	.02
Waiting for Child to Answer	-.14	-.23**	-.22*	-.15	.17	-.05	.00	-.09	-.02	-.02
Teacher Warmth	.07	-.13	.08	.06	.23*	.19*	.12	.12	-.03	.08
Teacher Honesty	.11	-.10	-.15	.03	-.06	.16	.40**	.18*	-.12	.23*

Table 26 (continued)

Correlation Coefficients Between Rating Scale Items and Factor Scores, and Spring Test Scores Partialling Out Fall Test Scores

	California Achievement Test Reading Comprehension		Reading Attitude Measure		California Achievement Test Mathematics Concepts		California Achievement Test Mathematics Computation		Mathematics Attitude Measure	
	2nd	5th	2nd	5th	2nd	5th	2nd	5th	2nd	5th
C.O.S. Factor I, Praise	.05	.09	.04	.14	.07	-.18	-.06	-.25	-.06	-.06
C.O.S. Factor II, Need for Intervention	-.15	.13	-.11	.04	-.31**	.13	.00	.21	.09	-.03
C.O.S. Factor III, Initiation of Activity	.07	-.04	-.22*	.01	-.23*	.06	.24*	.03	-.05	-.02

* $p \leq .10$

** $p \leq .05$

exercises in the fifth grade, the better the achievement in computation skills. This finding suggests that teacher control of the mathematics assignment is more desirable at the fifth grade level than at the second.

Factor III, labeled Initiation of Activity, indicates the extent to which students rather than teachers initiate most of the classroom interactions. In the second grade, this factor is negatively related to reading attitude and achievement in mathematics concepts, but positively related to achievement in mathematics computation. Classrooms where students initiate a number of contacts with the teachers probably are ones where the children check with teachers to learn if they have done an arithmetic problem right or have answered correctly in a reading workbook. Classes where this procedure is not followed are possibly ones where the teacher spends more time interacting with children teaching conceptual skills. Where teachers initiate questions and discussion about reading, the pupils come away with more of an appreciation for the subject.

Using the partial correlation technique, we were able to research for relationships which held across the entire range of teachers. We felt it was also important to determine if the videotape observation variables could be used to discriminate between teachers who had achieved large gains in pupil growth and between teachers who had not. To this end, the five teachers with the greatest gains as reflected by high intercepts and small slopes (Volume I, Chapter 7) were compared to the five teachers with the least gains (i.e., the "top" teachers and the "bottom" teachers) at

each grade level and in each subject area. Thus, the top five second grade reading teachers were compared with the bottom five second grade reading teachers on each variable, and so on. The non-parametric Mann-Whitney test was applied to the data to ferret out any significant differences. The mean ranks for each group from this test are recorded in Tables 27, 28 and 29.

None of the categories from the Teacher Instruction events discriminated between the two groups of teachers (Table 27). Only Weak Acknowledgement of Results from the Teacher Feedback events discriminated between the groups, with the effective teachers reacting more often with weak acknowledgement (Table 28).

The rating scale items did not significantly differentiate the two groups of teachers either. Teacher enthusiasm discriminated between second grade reading teachers, but the more effective teachers were not rated high in enthusiasm½ The variable of teacher enthusiasm must be studied very carefully in the future. The most effective second grade teachers were rated highest in having their Lessons Learning Centered. The only other significant relationship indicated that the most effective fifth grade mathematics teachers were rated highly in Teacher Honesty, a result consistent with the other findings on the rating of honesty.

The general lack of consistency between the top-bottom analysis and the partial correlation analysis is a bit surprising. This outcome may have occurred because the top and bottom teachers were identified on the basis of both slope and intercept rather than post-test achievement with pre-test partialled out. On the other hand, the result may also have come about because the top or bottom

five teachers truly are different from their colleagues and may tend to break the usual "rules" of good or poor teaching. More likely, however, the difference in results is due to the emphasis on gains per se in partial correlation and to the emphasis on patterns of gain in slope/intercept analysis. The reader is referred to Chapter 7 of Volume I for a further discussion.

Table 27

Mean Ranks in Comparisons Between Top and Bottom Reading and Mathematics Teachers on Teacher Instruction Lexicon Events

	Top Second Grade Reading (N = 5)	Bottom Second Grade Reading (N = 5)	Top Second Grade Math (N = 5)	Bottom Second Grade Math (N = 5)	Top Fifth Grade Reading (N = 5)	Bottom Fifth Grade Reading (N = 5)	Top Fifth Grade Math (N = 5)	Bottom Fifth Grade Math (N = 5)
Introductory Remarks	4.6	6.4	5.7	5.3	5.9	5.1	5.6	5.4
Expository Remarks	5.8	5.2	4.4	6.6	5.8	5.2	5.0	6.0
Giving Directions	5.3	5.7	6.0	5.0	5.8	5.2	4.6	6.4
Asking General Questions	7.0	4.0	5.6	5.4	5.9	5.1	5.7	5.3
Asking Prodding Questions or Cueing	6.3	4.7	5.4	5.6	5.3	5.7	5.3	5.7
Asking Questions to Determine if Pupil Has Understood Lesson	6.4	4.6	5.7	5.3	6.4	4.6	4.4	6.6

Table 28

Mean Ranks in Comparisons Between Top and Bottom Reading and Mathematics Teachers on Teacher Feedback Lexicon Events

	Top Second Grade Reading (N = 5)	Bottom Second Grade Reading (N = 5)	Top Second Grade Math (N = 5)	Bottom Second Grade Math (N = 5)	Top Fifth Grade Reading (N = 5)	Bottom Fifth Grade Reading (N = 5)	Top Fifth Grade Math (N = 5)	Bottom Fifth Grade Math (N = 5)
Simple Knowledge of Results	5.0	6.0	6.8	4.2	6.3	4.7	6.9	4.1
Elaborated Knowledge of Results	4.1	6.9	5.0	6.0	6.9	4.1	4.4	6.6
Child Asked to Elaborate	5.8	5.2	6.2	4.8	5.3	5.7	6.4	4.6
Positive Physical Contact	5.0	6.0	6.0	5.0	5.0	6.0	5.0	6.0
Encouraging Other Pupils to Give Feedback	6.0	5.0	5.4	5.6	5.0	6.0	5.6	5.4
Ignoring Positive or Mixed Achievement	6.7	4.3	4.4	6.6	5.4	5.6	6.9	4.1
Weak Acknowledgement of Achievement	7.2*	3.8	5.6	5.4	7.4**	3.6	4.0	7.0
Calling Attention to Error	6.0	5.0	5.2	5.8	5.0	6.0	5.7	5.3
Pupil Asked to Try Again	4.4	6.6	6.0	5.0	4.5	6.5	6.0	5.0
Praise for Good Behavior	5.8	5.2	4.2	6.8	6.7	4.3	4.5	6.5
Punishment for Disruptive Behavior	5.6	5.4	5.3	5.7	6.1	4.9	5.7	5.3
Warning for Negative Behavior	5.0	6.0	5.6	5.4	6.5	4.5	5.7	5.3

* $p \geq .10$

** $p \geq .05$

Table 29

Mean Ranks in Comparisons Between Top and Bottom Reading and Mathematics Teachers on Rating Scales

	Top Second Grade Reading (N = 5)	Bottom Second Grade Reading (N = 5)	Top Second Grade Math (N = 5)	Bottom Second Grade Math (N = 5)	Top Fifth Grade Reading (N = 5)	Bottom Fifth Grade Reading (N = 5)	Top Fifth Grade Math (N = 5)	Bottom Fifth Grade Math (N = 5)
Teacher Movement	4.9	6.1	5.2	5.8	5.6	5.4	4.6	6.4
Teacher Gestures	5.2	5.8	6.2	4.8	4.5	6.5	6.8	4.2
Teacher Enthusiasm	3.9*	7.1	5.6	5.4	5.9	5.1	6.0	5.0
Time Spent on Learning Activities	5.5	5.5	5.5	5.5	5.1	5.9	5.0	6.0
Lesson Learning Centered	7.2**	3.8	5.9	5.1	4.3	6.7	5.3	5.7
Differentiation of Learning Activities	5.4	5.6	6.0	5.0	6.6	4.4	6.7	4.3
Attentiveness of Children	5.0	6.0	6.0	5.0	6.0	5.0	6.1	4.9
Delay in Naming Child to Answer	6.0	5.0	6.8	4.2	5.9	5.1	5.0	6.0
Waiting for a Child to Answer	5.5	5.5	5.0	6.0	4.3	6.7	4.9	6.1
Teacher Warmth	5.4	5.6	6.3	4.7	5.5	6.0	5.5	5.5
Teacher Honesty	5.9	5.1	6.6	4.0	6.0	5.0	7.0*	4.0
Factor I, Praise	5.2	5.8	5.8	5.2	5.1	5.9	4.9	6.1
Factor II, Need for Intervention	5.7	5.3	5.6	5.4	4.8	6.2	5.2	5.8
Factor III, Initiation of Activity	6.4	4.6	6.4	4.6	5.8	5.2	5.3	5.7

* $p \geq .10$
** $p \geq .05$

Recommendations

The results presented in the preceding section support the notion that teacher behavior in classrooms can be studied by means of videotape, recording lessons as they occur in classrooms. All things considered, it is amazing that observations collected from viewing 40 minutes of one lesson during a year could be correlated with pupil achievement over the year's time. A short exposure to a teacher can provide information which will predict (to a small degree) pupil performance. The question of whether this is the best way to collect these data is not one that can be answered as of yet. Should future researchers choose to use the method employed in the present study, there are some recommendations we would like to bring forward concerning the recording of teacher behavior in naturalistic settings, the coding or observing of the teacher behavior once it has been videotaped, and specific hypotheses to test in future research.

Videotape recording. In videotape recording, observers found their job was easier when the sound was good and when the camera was focused on the teacher fairly closely, but also included some of the surrounding children. A portable, cordless microphone worn by the teacher seems to be best in recording teacher speech and should be considered in place of "shot-gun" types of microphones. We felt that the videorecording of what the pupils were doing was not very satisfactory. Therefore, we suggest that the cameraperson be informed about the behaviors which will later be coded, and concentrate on recording what the teacher does. The failure of RAMOS and the conventional APPLE system with the videotapes points out the limitations of the media for accurately recording detailed inter-

actions between specific children and teaches in a naturalistic setting. We did not find it possible to deal with all of the complexities of teacher-pupil relationships, for the camera could not get close enough without being obtrusive in some situations.

One option not explored in this study, which would keep down the costs of this technique and reduce any possible obtrusiveness, would be to have the teacher or pupils in the classroom make the videorecordings themselves. The machinery is easy enough to operate, and, with a small amount of training, teachers could record their own behavior thereby eliminating the need for a camera operator who is a stranger to the classroom. We suggest this procedure might be tested.

Videotape coding. We were not overly impressed with the interrater reliability that we obtained in the present study, although we were able to bring it up to an acceptable level. While we do not feel the observers in the present study were under-trained for their task, it is certain that an extensive amount of training for observers is highly important.

One possible problem in the present study was the large number of behaviors we asked the observers to attend to or look for. Future researchers may well consider requiring observers to attend to only a few categories of behaviors at a time. Rather than lexicons containing 25 categories, lexicons with only six or seven categories might yield more reliable results. This is not to say that the events observed need be global in nature, as with the ten-category Interaction Analysis system. Observers may look for quite discrete and specific behaviors, but should only look for a limited number at a time if accurate counts are desired.

Future researchers must also realize that there are various behaviors which are not very visible on videotapes. Some are not visible because they cannot be discriminated (e.g., an observer can not often tell if a teacher is smiling), others because they very rarely occur (see Tables 13-20).

Hypotheses for further research. One of the goals of Phase II of the BTES project was to generate hypotheses for further study. Several ideas have occurred to us, some general and some quite specific. In general, it is quite clear that teacher behaviors which are efficacious at one grade level are not necessarily effective at another grade level, and behaviors related to reading achievement and/or attitude do not necessarily have anything to do with behaviors related to mathematics achievement and/or attitude. For example, while Teacher Feedback categories correlated positively at the second grade, they correlated negatively at the fifth; while Asking General Questions was a Teacher Instruction behavior that was negatively related to mathematics, it was positively related to reading. Because teacher feedback is effective in the second grade and not at the fifth, perhaps social reinforcement from the teacher is much more important to second graders, and reinforcement for success at the activities themselves is more important at the fifth grade. One might then predict that the teacher as a "person" is important to second graders, and the teacher as a "planner" or "leader" is important to fifth graders.

The teacher as a person is also important at the fifth grade, as the results of the ratings of Teacher Warmth and Teacher Honesty illustrated. Research by Rogers and Carkhuff (Truax and Carkhuff, 1967) has indicated that these personality characteristics of counselors are

exceedingly important for growth in counsellees. Some of the variables from research on counselors, such as non-possessive warmth and positive regard, and honesty, might be further explored. Before doing so, however, it is important to realize that the teacher behaviors which are perceived as warmth, honesty, etc., are not behaviors which are easily modified or which may be taught in teacher training institutions. Candidates may be selected on the basis of these characteristics, but it may be unrealistic to try to train them to be "warm". However, before they can be used for selection, measures of these behaviors must be devised which have high reliability and high validity, and which are not culturally biased. The simple rating scales in the present study would have to be vastly improved. In fact, none of the measures used in the study had high enough correlations with achievement or discriminated well enough between top growth producing teachers and bottom growth producing teachers to justify their use in making decisions about individual teachers.

All of the significant results obtained and discussed in the previous section represent specific hypotheses to be tested and verified in more controlled settings. In interpreting the results, however, we have to bear in mind continually the fact that the videotapes only permitted the observation of teacher behavior. We could not tell to what extent teacher behavior was a cause or a result of achievement. For example, the teacher behavior of asking a child to try again after failure may be a good teaching practice, but asking children to try again does result from pupil failure. If we could control for the rate of pupil failure, then contrast the achievement in classrooms of teachers

who did require pupils to try again against those classrooms where teachers did not, we might be better able to evaluate this behavior.

All of the significant correlations in this study represent relationships; the specific cause and effect dynamic will have to be ferreted out by careful experimental design. We will not repeat the results as hypotheses; rather we refer the reader back to the section on the relationship between information collected from videotapes and pupil achievement.

Where categories of Teacher Instruction behavior or Teacher Feedback behavior were significantly related to achievement or attitude, it would be important to return to the data reduction section to determine which events were combined to make up the category. Perhaps these events can better specify an hypothesis.

The significant results from the rating scale data also suggest hypotheses. However, future researchers will have to operationalize the rating scales further, translating items into specific behaviors before proceeding with an experimental test. The ratings of teachers' question asking techniques, for example, deserve further study.

In summary, videotape recording does seem to have a place as a tool for studying teacher behavior. For some purposes, such as looking in detail at certain teacher behaviors, the advantages of observation through videotapes do outweight the disadvantages. No doubt this method will continue to be used in research studies of teacher behavior.

References

Flanders, N.A. The problems of observer training and reliability in Amidon, E.J., and Hough, J.B. (Eds.), Interaction Analysis: Theory, Research and Application. Reading, Mass: Addison-Wesley, 1967.

Fuller, F., and Manning, B.A. Self-confrontation reviewed: a conceptualization for video playback in teacher education. Review of Educational Research, 1973, 43, 469-528.

Medley, D.M., and Mitzel, H.E. Measuring classroom behavior by systematic observation in Gage, N.L. (Ed.) Handbook of Research on Teaching, Chicago: Rand McNally, 1963.

Truax, C.B., and Carkuff, R.R. Toward Effective Counseling and Psychotherapy: Training and Practice. Chicago: Aldine Publishing Co., 1967.

APPENDIX A

Definitions of Teacher Instruction Events

APPENDIX A

Definitions of Teacher Instruction Events

Event	Definition
1. Gaining Attention	A verbal statement or action which is intended to gain control of a child's attention prior to an instructional event.
2. Goal Statement	A statement of what is to be learned from the lesson; the objectives or outcomes of an instructional activity.
3. Stimulating Recall of Prerequisites	A direction to remember a previous skill or information earlier learned; the purpose is to tie it to the present lesson.
4. Explanations	
4A. Statement of fact	The teacher's initiation of a statement to be learned as an abstraction, idea, fact, definition, generalization, principle, rule, etc.
4A1. On the topic (when the fact is revelant to the lesson)	
4A2. Off the topic (when the fact is irrevelant)	
4B. Examples	The teacher gives an instance or illustration which represents a concept previously explicated.
4C. Demonstration and explanation	The teacher gives a physical demonstration illustrating a concept or principle previously explicated.

APPENDIX A

Definition of Teacher Instruction Events

Event	Definition
4. Explanations (continued)	
4C1. On topic	
4C2. Off topic	
4D. Directions	Directions given by the teacher designed to create an active response on the part of the pupil or pupils.
4D1. On topic	
4D2. Off topic	
4E. Summarizing	A statement bringing together several previous statements so as to present a uniform whole.
4F. Giving applications	The teacher mentions how what has been learned might be applied to a new situation not practiced in class.
4G. Contrasting and comparing	Giving similarities and differences in a number of concepts - either concepts in the same lesson or contrasting previously learned concepts with newly learned ones.
4H. Cueing	Supplying part or all of an answer to stimulate the student to complete the answer. "Giving hints" in pursuit of a response from the child.

APPENDIX A

Definitions of Teacher Instruction Events

Event	Definition
5. Questioning for Active Response	Asking a question for a factual answer which is open (asking for multiple answers or divergent thinking) and which is geared at applying or synthesizing information (a high cognitive level).
5A. Asking for answer	Asking a question for a factual answer which is open and which calls for recall of information or identification.
	Asking a question for a factual answer which calls for a single right answer (convergent thinking - closed question) and which calls for an application, synthesis or other high level thinking.
	Asking a question for a factual answer which calls for a single right answer and for simple recall or identification.
5BA. Asking a follow-up question - for addition information	The teacher asks the child or group to supply some more information in addition to that already given. The implication is that the answer is not complete.
5BD. Asking a follow-up question - for development of response	The teacher asks the child or group to <u>expand</u> <u>or</u> <u>elaborate</u> upon an answer already given, but not necessarily give new information.

APPENDIX A

Definitions of Teacher Instruction Events

Event	Definition
5. Questioning for Active Response (continued)	
5BP. Asking a follow-up question - for application of response	The teacher asks for a transfer or application of what has been said to a new situation.
5C. Asking for agreement	Asking if the class or others in the class agree with what has been said.
5D. Asking for disagreement	Asking if the class members disagree with what has been said.
5E. Asking for judgement	Asking the class to evaluate or judge something that has been said or presented, either as a follow-up or as an initial question.
5F. Asking for a performance appraisal	Asking a question to determine if a child has mastered a lesson and to determine his status in learning; testing.
5G. Asking if understood	The teacher asks for confirmation that what has been said has been heard or understood; seeking feedback in regard to presentation of discussion.

APPENDIX B

Definitions of Teacher Feedback Events

APPENDIX B

Definitions of Teacher Feedback Events

Event	Definition
I. Teacher Reactions to Positive Achievement or Behavior	
1A. Simple knowledge of results	The unelaborated, usually one word, indication that the behavior or response is correct.
1B. Elaborated knowledge of results	An explanation of the reasoning behind an answer or behavior, an elaboration or improvement of the answer given by the teacher.
1C. Direction to child to elaborate	The direction to elaborate, extend, improve, or add to answers along with the indication (verbal or nonverbal) that the answer is correct.
1D1. General verbal praise	A general statement of praise for the pupil because he has responded appropriately.
1D2. Granting privileges	Allowing the pupil to undertake a desired activity because he has performed well.
1D3. Positive physical contact	Giving a physical display of affection to child following good work or behavior.

APPENDIX B

Definitions of Teacher Feedback Events

Event	Definition
I. Teacher Reaction to Positive Achievement or Behavior (continued)	
1D4. Award of consumables, tokens, or symbols of recognition	Giving a tangible object or an abstract symbol to a child following good work or behavior.
1E. Teacher directs other to give feedback	While acknowledging correctness of the response, the teacher asks other members of the class to admire, judge or add to the response.
1F. Ignoring of positive behavior or achievement	When the teacher does not pay attention to a correct response that seems to have been perceived; seemingly aware but not responding.
II. Teacher Noncommittal Reactions to Any Achievement or Behavior	
2A. Acknowledges perceiving response	When the teacher indicates the response has been perceived but gives no indication of any evaluation.
2B1. Indication of both positive and negative elements in response, both weighed equally	When the teacher indicates that he is both pleased and displeased with a response.

APPENDIX B

Definitions of Teacher Feedback Events

Event	Definition
II. Teacher Noncommittal Reactions to Any Achievement or Behavior (continued)	
2B2. Indication of both positive and negative elements in response, emphasis on the positive	The teacher indicates he is primarily happy with the response although it is not quite correct.
2B3. Indication of both positive and negative elements in response, emphasis on the negative	The teacher acknowledges that part of a response is correct or in the right direction but that the response is primarily incorrect.
2C. Direction to child to elaborate	The direction to elaborate, extend, improve, or add to answers without any indication (verbal or nonverbal) that the answer is correct.
2D. Promise of reward contingent on change or continuation	Without indicating if the presently engaged in behavior is right or wrong, the teacher specifies the how a reward may be achieved.
2E. Directs others to give feedback	While giving no information about the correctness of the response, the teacher asks other members of the class to admire, judge, change or add to the response.

APPENDIX B

Definitions of Teacher Feedback Events

Event	Definition
II. Teacher Noncommittal Reactions to Any Achievement or Behavior	
(continued)	
2F. Ignores part of response	The ignoring of one aspect of a pupil's response while concentrating on another aspect.
III. Teacher Reactions to Negative Achievement or Behavior	
3A1. Simple knowledge of results-response termed incorrect	The unelaborated, usually one word, indication that the response is incorrect or wrong.
3A2. Simple knowledge of results - correct answer given	The unelaborated correct answer is supplied to child in error.
3B1. Error elaborated by teacher	An explanation, discussion or elaboration of the reasoning behind an error or wrong thinking.
3B2. Correct answer or behavior elaborated by teacher	When the teacher supplies the reasons or thinking behind the desired correct response.
3C. Direction to try again	When the teacher asks the child to make a new and different response after failure or partial success, or asks child to evaluate his own responses.

APPENDIX B

Definitions of Teacher Feedback Events

Event	Definition
III. Teacher Reactions to Negative Achievement or Behavior (continued)	
3D. General punishment	
3D1. Verbal reprimand	A general statement of criticism directed at the pupil or class because they have not responded appropriately; often consists of sarcasm, or guilt inducing statements.
3D2. Withdrawal of privileges	The child or group is not allowed to undertake a desired activity that is scheduled or has been promised because of negative behaviors. Includes the assignment of extra tasks.
3D3. Aversive physical contact	Physical contact with the child designed to cause him pain or embarrassment. Corporal punishment.
3D4. Denial of consumables, tokens or symbols or recognition	Activity denying a tangible object or symbol to child or children following poor work or behavior.
3D5. Directions to stop activity	When the child or children are told to stop doing what they are presently engaged in <u>because</u> it is incorrect or wrong.

APPENDIX B

Definitions of Teacher Feedback Events

Event	Definition
III. Teacher Reactions to Negative Achievement or Behavior (continued)	
3D6. Warning of consequences	The teacher informs the child what the consequences will be for him if he continues to act in the same fashion.
3D7. Directions to proceed with assigned activity	Instructions for child or group to redirect their attention to the assigned task, with or without a threat.
3D8. Isolation, time out or sent from room	As a result of some action, one or several individuals is removed from the ongoing activities of the class.
3E. Directs others to give feedback	While indicating the incorrectness of a response or action, the teacher asks the other members of the class to criticize or to state the desired answer or action.
3F. Ignoring negative behavior	When the teacher obviously does not pay attention to an incorrect response that seems to have been perceived; seemingly award but not responding.
3G. Teacher acknowledges defeat	When the teacher indicates that he or she does not know what to do in response to negative behavior, and gives up.

APPENDIX B

Definitions of Teacher Feedback Events

Event	Definition
III. Teacher Reactions to Negative Achievement or Behavior (continued)	
3H1. Teacher acceptance of pupil's lack of answer	When the pupil indicates he does not know an answer or procedure, the teacher does not reprimand him.
3H2. Teacher rejection of pupil's lack of answer	When the pupil does not answer or says "I don't know" he is chided.

APPENDIX C

Observation Classroom Summary Record: APPLE System

Videotape Rating Schedule #1

Videotape Rating Schedule #2

OBSERVATION
CLASSROOM SUMMARY RECORD:
APPLE SYSTEM

Observer _____ Date _____

Teacher _____ Grade _____ School _____

1. There was frequent praise by an adult.

4	3	2	1
very descriptive			not at all descriptive

2. Praise was directed at many different students.

4	3	2	1
very descriptive			not at all descriptive

3. Students initiated most adult-student interactions.

4	3	2	1
very descriptive			not at all descriptive

4. Adults initiated most adult-student interactions.

4	3	2	1
very descriptive			not at all descriptive

5. Students gave a lot of responses.

4	3	2	1
very descriptive			not at all descriptive

6. Student responses were mostly correct.

4	3	2	1
very descriptive			not at all descriptive

7. Student errors were usually followed by corrective feedback.

4	3	2	1
very descriptive			not at all descriptive

8. After the first half hour, most students disregarded the observer.

4	3	2	1
very descriptive			not at all descriptive

9. Adults organized most groups.

4	3	2	1
very descriptive			not at all descriptive

10. Students worked in previously organized groups.

4	3	2	1
very descriptive			not at all descriptive

11. The entire class worked on the same subject at the same time.

4	3	2	1
very descriptive			not at all descriptive

12. Individual students usually decided how they were going to spend their day.

4	3	2	1
very descriptive			not at all descriptive

13. Students were generally in control of themselves.

4	3	2	1
very descriptive			not at all descriptive

14. There was often a need for adult intervention.

4	3	2	1
very descriptive			not at all descriptive

15. Adults instructed students for most of the instructional day.

4	3	2	1
very descriptive			not at all descriptive

16. Students worked independently for most of the instructional day.

4	3	2	1
very descriptive			not at all descriptive

17. Students spent most of day in groups of 5 or more.

4	3	2	1
very descriptive			not at all descriptive

18. Students spent most of day in groups of 4 or less.

4	3	2	1
very descriptive			not at all descriptive

19. Most students did their work in reading and math without needing to be reminded, rewarded, or otherwise specifically motivated by an adult.

4	3	2	1
very descriptive			not at all descriptive

20. Most of the students required considerable specific motivation by an adult to do their work in reading and math.

4	3	2	1
very descriptive			not at all descriptive

21. Most students were generally friendly toward each other during the day.

4	3	2	1
very descriptive			not at all descriptive

22. There was occasional unfriendliness between students throughout the day.

4	3	2	1
very descriptive			not at all descriptive

23. There was evidence of a systematic effort to monitor the progress of every pupil on assigned tasks.

4	3	2	1
very descriptive			not at all descriptive

24. Adults frequently found the use of negative reinforcement necessary.

4	3	2	1
very descriptive			not at all descriptive

Comments:

_____

_____

_____

Rater_____
Teacher_____
Topic_____
Date recorded_____

VIDEOTAPE RATING SCHEDULE #1

1. Variability During Lesson

☐ Large number of different activities; shifts in topic; alternation of principle and example constantly.	☐ Several different activities during lesson; many shifts in topic; frequent alternation between principle and example.	☐ During the lesson there are one or two activities; a few shifts in topic and some alternation of principle and example.	☐ There is only one activity for most of the pupils; only one topic or so it seems.

2. Pace of Lesson

☐ Fast	☐ Lively	☐ Leisurely	☐ Slow

3. Tempo of Lesson

☐ Pace varies considerably. Some parts slow, some parts fast.	☐ Pace varies quite a bit from slow to fast.	☐ Pace is predominantly fast and sometimes slow or predominantly slow and sometimes fast. Some change in tempo.	☐ The pace is uniformly slow or uniformly fast and the tempo does not change much from time to time.

4. Teacher Movement

☐ Teacher moves purposefully about the class constantly for different activities.	☐ Teacher moves from position to position quite often.	☐ Teacher stays in one position for most of the lesson and often does not move appropriately.	☐ Teacher is relatively immobile and moves very little during the lesson. Maintains one position.

Copyright © 1974 by Educational Testing Service. All rights reserved.

5. Teacher Gestures

☐	☐	☐	☐
Teacher constantly uses hands or makes body movements to emphasize what is being said.	Teacher is very prone to use hand gestures and other body movements to make a point.	Teacher sometimes uses gestures to make a point but most of the time is immobile.	Teacher seldom or never gesticulates. Very wooden in delivery style.

6. Pupil Enjoyment

☐	☐	☐	☐
Pupils seem very happy and at ease.	Pupils are happy much of the time.	Pupils more neutral or sad than happy.	Pupils never display enjoyment or pleasure in learning.

7. Teacher Enthusiasm

☐	☐	☐	☐
Teacher is obviously excited about the topic--is alert, stimulating and endearing.	Teacher is fairly enthusiastic about the topic.	Teacher is moderately interested in the subject.	Teacher seems tired of topic and reacts in patterned, routine manner.

8. Structuring of Learning Environment

☐	☐	☐	☐
Activities are completely structured for the pupils. Teacher is in complete control of the learning environment through direct or indirect methods.	Considerable structure present. Teacher directs the child learning fairly closely.	Some structure evident. The teacher has loose control of child learning.	No structure observable. Teacher not in control of child learning either directly or indirectly.

9. Time Spent on Learning Activities

☐	☐	☐	☐
All of the time in lesson spent on learning activities.	Most of the time spent on learning; some effort spent on management.	Some time spent on management; some time spent on learning.	All of the time is spent on management or other activities.

10. Teacher Organization

☐	☐	☐	☐
Lesson superbly organized and obviously pre-planned.	Lesson is organized and some planning is evident.	Lesson somewhat disorganized with the appearance of minimal planning.	Lesson completely disorganized. Teacher seems to be improvising on the spot.

11. Lesson Learning Centered

☐	☐	☐	☐
Lesson emphasizes the learning of cognitive skills or knowledge.	Main emphasis is on cognitive learning with some emphasis on having a pleasurable time.	About half the time spent on learning, half the time spent on enjoyment of the lesson.	Lesson centered on enjoying time together--having a good time.

12. Differentiation of Learning Activities

☐	☐	☐	☐
All or majority of pupils doing similar assignments.	Large groups of pupils with similar materials.	Several small groups working on different assignments.	Each child seems to have his own individual assignment.

13. Teacher Sensitivity to Feelings

☐	☐	☐	☐
Highly sensitive, able to reflect feelings back to the child or otherwise communicate empathy.	Sensitive and aware much of the time to the pupil's emotional experiences in the classroom.	Not very sensitive to pupils in distress or having other emotional experiences.	Relatively insensitive to pupil's emotional experiences in the classroom. Little empathy sensed.

14. Body Language Indicating Emotional Support

☐	☐	☐	☐
Consistently moves toward child assuming a posture indicating support to those expressing affect-laden information. Highly communicative and appropriate facial expressions.	Often attentive to child's emotional needs. Gives appropriate support much of the time with body movement and facial expression.	Physical position and attitude of body along with facial expression indicates support only some of the time when a pupil evidences emotional needs.	Almost no communication of support by facial expressions or body positions and postures. No non-verbal acknowledgement of emotional needs.

Teacher Code _____

Tape Topic _____

VIDEOTAPE RATING SCHEDULE #2

Tape Date (Taping) _____

Rater _____

1. How attentive are the children at whom the instruction is aimed?

☐	☐	☐	☐
Always attentive (all of them)	Attentive most of the time (most of them)	Attentive sometimes (some of them)	Seldom attentive (few of them)

2. When does the teacher name the child to answer when asking questions?

☐	☐	☐	☐
Always names a child to answer first before asking a question	Usually calls a name first before asking a question	Usually asks the question first before calling a name	Always asks a question first before calling on a particular child to answer

3. How long does the teacher wait for a reply after asking a question?

☐	☐	☐	☐
Waits only a fraction of a second	Waits a second or two or until answers	Waits until clear a child does not know the answer (3-5 sec.)	Waits for child to the point of embarrassment

4. Teacher warmth.

☐	☐	☐	☐
Very warm and accepting of pupils	Warm but sometimes cool with pupils	Cool but sometimes warm with pupils	Cool to students and unaccepting of pupil concerns or problems

5. Teacher honesty.

☐	☐	☐	☐
Very honest with children/ Completely open about both positive and negative feelings	Honest with children most of the time/ Sometimes does or says things without meaning them	Honest with pupils sometimes, sometimes not	Dishonest with children/ Does not display genuine feelings or opinions

Copyright © 1974 by Educational Testing Service. All rights reserved.